Scientific Report

on the ESF Exploratory Workshop

Understanding the Dynamics of Knowledge

Certosa di Pontignano, Siena, Italy

17-19 November 2005

Chairs:

Cristiano Castelfranchi and Boicho Kokinov

Local organizer (contact person):

Fabio Paglieri

paglieri@media.unisi.it

CONTENTS

1. Executive summary

As stated in the original proposal, this interdisciplinary Exploratory Workshop aimed to integrate various aspects of knowledge dynamics in human and artificial agents, including evolution of concepts, cognitive development and learning, and short-term dynamics such as belief change and information update. Both individual and social dynamics of knowledge were covered, and their interplay analyzed. Formal and computational models were compared with socio-cognitive theories of knowledge change, and with empirical findings in psychology, anthropology, and social sciences. The workshop proved to be extremely successful, fulfilling its scientific objectives and fostering future research cooperation among the participants and their institutions (see section 3 for details), mainly due to the lively and insightful debate that resulted from the meeting, both during the formal sessions and in informal gatherings (breaks, meals, etc.). A total of 34 scholars from 11 countries participated, 25 of them as invited speakers, the remaining 9 as registered attendees (see section 5 and 6 for further details).
The final programme and the abstract of each talk are provided, respectively, in section 4 and 2. Here a short survey of the main thematic axes of the event is given, as they come to be discussed during the workshop.

SHORT-TERM, LONG-TERM AND EVOLUTIONARY DYNAMICS OF KNOWLEDGE
Different dynamics of knowledge can be classified according to their time-scale: *short-term dynamics* (e.g. context effects, priming, belief revision strategies), *long-term dynamics* (e.g. learning and development), *evolutionary dynamics* (e.g. genetic heritage and cultural change). Although often fruitful in suggesting the most adequate approach to specific phenomena of knowledge change (like the use of epidemiological models to understand cultural transmission), this categorization cannot remain rigid and static. All these different layers of knowledge continuously interact with each other: short-term dynamics evolves dramatically during development (e.g. false belief attribution) and it is also affected by learning and cultural variation (e.g. default trust attribution to specific information channels is a direct effect of cultural values and educational practices); similarly, long-term dynamics like learning and development needs to be integrated into (and justified by) broader frameworks, respectively the socio-cultural context of individuals and the evolutionary history of species; finally, evolutionary explanations of knowledge dynamics must prove consistent with empirical evidences from the other levels – explaining why and how humans have evolved certain specific short-term and long-term dynamics rather than others, and what needs, pressures and preconditions have been answered by such dynamics. The first part of the Workshop (see section 4) was especially devoted to explore this family of problems, fostering closer comparison and future integration between theories of knowledge dynamics at various time-scales.

INDIVIDUAL AND SOCIAL DYNAMICS OF KNOWLEDGE
Individual and social dynamics of knowledge cannot be understood separately from each other: the social creation, transmission and distribution of knowledge is an emergent phenomenon from the complex interaction of individual agents; at the same time, the cognitive processes of knowledge acquisition, interpretation, elaboration and generation are shaped and motivated by several social factors, e.g. cultural values, norms, reputation, communication protocols. How does social dynamics of knowledge emerge from individual processes of information update, belief revision, inter-personal argumentation? How these cognitive processes are in turn influenced and partially shaped by social pressures, with special reference to knowledge change, propagation and availability? Which paradigms and tools are most appropriate to simulate in artificial societies both the social emergence of large-scale knowledge dynamics, and their feedbacks on small groups and individual cognition? The second day of the Workshop was mainly focused on such complex interactions between micro- and macro-level of knowledge dynamics (see section 4).

SOCIO-COGNITIVE THEORIES AND FORMAL AND COMPUTATIONAL MODELS OF KNOWLEDGE CHANGE, DEVELOPMENT AND EVOLUTION

Several aspects of knowledge change, like belief revision, information update and argumentation, have been extensively studied within strong *formal and computational frameworks*, both in logic, Artificial Intelligence and cognitive economics. While these approaches provide powerful tools of formalization and promising opportunities for practical applications (e.g. information retrieval, belief-based decision making, trust evaluation for security), they would in turn greatly profit from *more in-depth cross-fertilization with other socio-cognitive disciplines*, like cognitive psychology, social sciences, developmental studies, linguistic, evolutionary and comparative anthropology. This interchange would broaden the perspective of current formalisms, defining new relevant features of knowledge dynamics to be modelled and reproduced in artificial systems; on the other hand, socio-cognitive theories of knowledge change would equally benefit from updated logical and computational models, being able to test their predictions within better simulative frameworks (e.g. connectionist models for Artificial Life, multi-agent social simulations) and to study the interplay of more sophisticated and realistic artificial cognitive agents (e.g. fully autonomous, goal-oriented, belief-based agents). To this purpose, the third part of the Workshop brought together world-leading experts in logic, cognitive economics, computer science and Artificial Intelligence, to face the challenge of outlining new formal and computational paradigms for human knowledge and its dynamics.

2. Scientific content of the event

The main scientific contents of the workshop are summarized in the abstracts below: they are organized following the structure and the chronological order of the event itself. All the authors gave permission to distribute their abstracts and their e-mail addresses as part of this Scientific Report.

Thursday 17 November 2005, 9:00-13:00, first session:

2.1. Short-term dynamics of knowledge: Cognitive and computational models of belief change

Chair: Elizabeth Robinson (University of Warwick)

Context-sensitivity of human cognition: Fast short-term restructuring and adaptation of the cognitive system based on what is anticipated to be relevant

Boicho Kokinov

New Bulgarian University, bkokinov@nbu.bg

Learning produces long-term changes in human beliefs, concepts, and skills. However, even without learning the human cognitive system is subject to short-term changes that produce variability in human behavior. Thus, even without learning new facts, one can change the decisions already made, one can fail solving a problem that was previously solved, one can judge the same option differently. Why is that? Why is human behavior so unstable? Some researchers would claim that this is random noise due to the non-deterministic machinery of human cognition. However, our claim is that human behavior changes systematically to adapt to the changes in the environment.

This talk presents experimental material demonstrating context effects on various cognitive processes, including problem solving, decision-making, and judgment. It is demonstrated how small changes in supposedly irrelevant elements of the environment can change the outcome of a cognitive process without the subject to be aware of that fact. A series of experiments is reviewed which shows that people change the way they solve a problem (Kokinov & Yoveva, 1996, Kokinov, Hadjiilieva, Yoveva, 1997), the judgment they make of an object – a line, an age, or a price (Kokinov, Hristova, Petkov, 2004, Hristova, Petkov, Kokinov, 2005), or even their willingness to risk in a decision-making task (Kokinov, Raeva, 2004).

These phenomena are accounted for within the general cognitive architecture DUAL and simulated with a series of models built on it – AMBR (Kokinov, Grinberg, 2001) and JUDGEMAP (Kokinov, Hristova, Petkov, 2004). It is argued that human cognitive system is being restructured and various subsets of concepts, beliefs, and skills become available with every change of context. In such a way human cognitive system adapts to the environment and becomes more effective since it anticipates certain aspects of that environment to be relevant while others – not to be relevant. This allows the cognitive system to re-represent the environment, the task, and its own beliefs in a context-relevant way, to access only the relevant pieces of knowledge, to involve only the relevant mechanisms in the computations and effectively to find a solution to the problem. The interaction between bottom-up pressure (from perception) and top-down pressure (from goals) in computing the relevance is discussed.

When we face a task that is atypical for the particular environment our anticipation of what is relevant may fail and we may fail in solving the problem or make a strange decision. However, in most cases we face the tasks in their typical environments (the environments in which they typically occur) and the mechanisms for relevance anticipation described above make it possible for effective decision-making and problem solving and eventually for fast reaction. Animals who live in environments that are relatively stable in terms of what is relevant for the creature can be less flexible and rely on fixed instincts and mechanisms, however, human beings live in environments that can change radically within seconds and where the goals change accordingly, and therefore what was considered relevant a minute ago is no longer relevant. Thus context-sensitivity is a survival mechanism in dynamic environments that are changing dramatically within minutes.

REFERENCES

Hristova, P., Petkov, G., Kokinov, B. (2005). "Influence of Irrelevant Information on Price Judgments". In: *Advances in Cognitive Economics*. Sofia: NBU Press

Kokinov, B., Raeva, D. (2004). "Can an Incidental Picture Make Us More or Less Willing to Risk?" *Proceedings of the 1st European Conference on Cognitive Economics*.

Kokinov, B., Hristova, P., Petkov, G. (2004). "Does Irrelevant Information Play a Role in Judgment?" In: *Proceedings of the 26th Annual Conference of the Cognitive Science Society*. Erlbaum, Hillsdale, NJ.

Kokinov, B., Grinberg, M. (2001). "Simulating Context Effects in Problem Solving with AMBR". In: Akman, V., Thomason, R., Bouquet, P. (eds.) *Modeling and Using Context*. Lecture Notes in Computer Science (Lecture Notes in Artificial Intelligence), vol. 1775, Springer Verlag.

Kokinov, B. (1999). "Dynamics and Automaticity of Context: A Cognitive Modelling Approach". In: Bouquet, P., Serafini, L., Brezillon, P., Benerecetti, M., Castellani, F. (eds.) *Modeling and Using Context*. Lecture Notes in Computer Science (Lecture Notes in Artificial Intelligence), vol. 1688, Springer Verlag.

Kokinov, B., Hadjiilieva, K., Yoveva, M. (1997). "Explicit vs. Imlicit Hint: Which One is More Useful?" In: Kokinov, B. (ed.) – *Perspectives on Cognitive Science*. Vol. 3. NBU Press, Sofia.

Kokinov, B. (1997). "A Dynamic Theory of Implicit Context". In: *Proceedings of the 2nd European Conference on Cognitive Science*. Univ. of Manchester Press.

Kokinov, B., Yoveva, M. (1996). "Context Effects on Problem Solving". In: *Proceedings of the 18th Annual Conference of the Cognitive Science Society*. Erlbaum, Hillsdale, NJ.

Fluidly represent the world: Way, way harder than you think

Robert French

LEAD-CNRS & University of Burgundy at Djion, robert.french@u-bourgogne.fr

We will attempt to show that representing objects, situations, and actions in the world is *much* harder than one might suspect. We will point to some of the problems that underestimating the difficulties involved in representation-building caused for traditional Artificial Intelligence. We believe that representations must ultimately arise from a dynamical exchange between top-down and bottom-up processes, and I will defend this point of view during this talk (and throughout the Workshop). We will discuss the evolution of representations, from the purely perception-based (i.e., bottom-up) representations of three- to four-month old infants to perceptual/conceptual representations of adults. We will show that things start to get interesting when concepts begin to become part of long-term memory and begin to dynamically interact with bottom-up perceptual processes. We examine, largely through examples drawn from a number of different domains, some of the severe problems caused by assuming representational and conceptual fixity.

From knowledge to action: Reason-based belief dynamics and belief-based goal dynamics*

Cristiano Castelfranchi

ISTC-CNR Roma & University of Siena, cristiano.castelfranchi@istc.cnr.it

*An extended version of this abstract is available on line at http://www.media.unisi.it/cirg/udk/abs_esf_castelfranchi.doc

In this talk I will critically discuss two crucial features for cognitive models and formal theories of belief dynamics: the need to develop a better understanding of the *reasons by which a rational agent comes to believe* something; and the deep connections and rich interaction between *belief dynamics* and *goal dynamics*.

As for the first topic, it is a fact of life that we cannot believe everything we observe or that we are told. We accept a given information or datum as a belief on the basis of our previous beliefs, of its evidences, supports and sources, and of others psychological factors. Here I will sketch some crucial points of these cognitive mechanisms.

My fundamental claim is that our knowledge base is not (and should not be modelled as) a file where one can introduce new data or eliminate a file-card without altering the other data. Our beliefs are integrated, interconnected and mutually supported: to drop a belief or to add a new one entails checking its coherence with other beliefs and revising previous knowledge. The belief-belief coherence and support is quite a well studied problem in philosophy and AI (truth maintenance systems; belief revision and updating; argumentation) and in some cognitive agent architectures. There are in fact two schools in belief revision (Harman, 1986; Gärdenfors, 1988; Doyle, 1992): the foundations approach stressing the importance of supports and justifications of beliefs, and the coherence approach modelling logical compatibility and coherence. However, I agree with Doyle (1992) that there is no incompatibility between the two models, and that rational beliefs must be both justified and relatively coherent.

In this light, I will present in the first part of my talk several basic features of a *reason-based theory of belief formation and change*: among other topics, I will discuss the distinction between storing and believing a given information (Cantewell, 1996; Castelfranchi, 1997; Paglieri, 2004), the real nature of the so called ‘decision to believe’ (Harman, 1986; Castelfranchi, 1996), the main characteristics of source reliability (Fullam, 2003; Falcone, Castelfranchi, 2004) and belief credibility (Castelfranchi, 1996; 1997; Paglieri, 2004), and the basic patterns of motivational influence over belief formation and change (Frijda et al., 2000; Paglieri, 2005).

As for the interaction between goal dynamics and belief dynamics, I will concentrate on the issue of intentions adoption and revision, and of its strong relation with belief formation and change. In general, what Bratman (1990) calls “coherence”, and Cohen and Levesque's (1990) “rational equilibrium” between the agent's intentions and beliefs, is reduced only to the fact that the agent selects and adopts those intentions that he believes to be achievable. In current BDI models, beliefs are of course crucial for the adoption or the abandoning of intentions, but their role seems quite limited: during the processing the belief component is not consulted at each step (consider for instance Rao and Georgeff’s (1991) architecture): some crucial steps, like planning, are not based on beliefs (means-end and causal relations). Only in Bratman, Israel and Pollack’s architecture (1988), beliefs enter all the components of the architecture, determining activation, deliberation, planning, etc. In some sense, I will make explicit such a role of beliefs in the process, adding also the idea of their supporting role, and of their effect on the "quality" of the goal.

In fact, in those models so far there is no clear distinction among :

- the *Processing of goals***,** from their firing to their satisfaction or abandon: how beliefs determine such a process step by step;
- the *Dynamics or Revision of goals***,** i.e. the change of goals ("motivations", "preferences", "desires", depending on the terminology of different authors) on the basis of changes in a dynamic external environment, or internal cycles of the agent;
- the *Typology of goals***,** that may be partially characterized just on the basis of their typical *belief structure*.

Of course, there are relations among these different aspects of goal theory in which *belief structure* is relevant. Normally, the processing of a goal from its firing to its satisfaction is intertwined with the Dynamics of goals (changing goal, or the activation of other goals, etc.). Also the differences among kinds of goals (like "intentions" vs. "desires", or "expectations" vs. "renounces", etc.) are frequently related to different steps in the goal processing.

A general theory of this relation is needed, that should include, in my view, four claims about the role of beliefs relative to goals' life:

- beliefs support goals (they become their Reasons);
- beliefs determine goal processing;
- beliefs determine goal dynamics (revision);
- beliefs determine goal kinds.

We maintain in our mind both *reasons to believe*, and *reasons to do.* We need to have "reasons" both for believing and for aiming at something. We cannot do this arbitrarily. This is the *common feature of both faces of our "rationality"*: belief rationality (epistemic) and goal rationality (pragmatic). "Reasons" give the agent the possibility to *justify* and *explain* (to itself and to others) its actions, being in this way a major aspect of its *rationality* and of its consciousness. The second part of my talk will be devoted to explore and discuss in detail those issues.

Finally, I will conclude offering a couple of tentative speculations: first, I will suggest that a reason-based account of belief and goal dynamics foster our understanding of the crucial notion of *agent autonomy* (Castelfranchi, 1995), i.e. those cognitive and behavioural capabilities that an intelligent agent must show, in order to considered (to some extent) autonomous; second, I will shortly speculate on the *need for coherence* that the human mind shows both in belief dynamics and in goal processing (Paglieri, Castelfranchi, 2005), and on what we may learn from such common pattern in epistemic and pragmatic rationality.

REFERENCES

Bratman, M.E. 1990. “What is an Intention?”. In *Intentions in Communication*, P.R Cohen, J. Morgan, M.A. Pollack (eds.), pp.15-32. Cambridge, Mass.: MIT Press.

Bratman, M.E., Israel, D.J., Pollack, M.E. 1988. “Plans and resource-bounded practical reasoning”. *Computational Intelligence* **4**, pp. 349-355.

Cantwell, J. 1996. *Resolving Conflicting Information*. Technical Report, Department of Philosophy, Uppsala University, June 20, 1996.

Castelfranchi, C. 1995. “Guarantees for Autonomy in Cognitive Agent Architecture”. In *Proceedings ECAI 1994 Workshop on Agent Theories, Architectures, and Languages*, pp. 56-70. Berlin: Springer.

Castelfranchi, C. 1996. “Reasons: Belief Support and Goal Dynamics”. *Mathware & Soft Computing* **3**, pp. 233-247.

Castelfranchi, C. 1997. “Representation and Integration of Multiple Knowledge Sources: Issues and Questions”. In *Human & Machine Perception: Information Fusion*, Cantoni, Di Gesù, Setti, Tegolo (eds.), Plenum Press.

Cohen, P. R., Levesque, H.J. 1990. “Rational Interaction as the Basis for Communication”. In *Intentions in Communication*, P.R Cohen, J. Morgan, M.A. Pollack (eds.), pp. 33-71. Cambridge, Mass.: MIT Press.

Doyle, J. 1992. “Reason Maintenance and Belief Revision: Foundations vs. Coherence Theories”. In *Belief Revision*, P. Gärdenfors (ed.), pp. 29-51. Cambridge: Cambridge University Press.

Falcone, R., Castelfranchi, C. 2004. “Trust Dynamics: How Trust Is Influenced by Direct Experiences and by Trust Itself”. In *Proceedings of AAMAS 2004*, N.R. Jennings, C. Sierra, L. Sonenberg, M. Tambe (eds.), pp. 740-747. New York: ACM Press.

Frijda, N. H., Manstead, A., Bem, S. (eds.) 2000. *Emotions and Beliefs: How Feelings Influence Thoughts*. Cambridge: Cambridge University Press.

Fullam, K. 2003. *An Expressive Belief Revision Framework Based on Information Valuation*. MS thesis, University of Texas at Austin.

Gärdenfors, P. 1988. *Knowledge in Flux: Modeling the Dynamics of Epistemic States*. Cambridge, Mass.: MIT Press.

Gilbert, D.T. 1991. “How Mental Systems Believe”. *American Psychologist* **46**, pp. 107-119.

Harman, G. 1986. *Changes in View: Principles of Reasoning*. Cambridge, Mass.: MIT Press.

Paglieri, F. 2004. “Data-oriented Belief Revision: Towards a Unified Theory of Epistemic Processing”. In E. Onaindia, S. Staab (eds.), *Proceeding of STAIRS 2004*. Amsterdam: IOS Press.

Paglieri, F., Castelfranchi, C. 2005. “Revising Beliefs Through Arguments: Bridging the Gap between Argumentation and Belief Revision in MAS”. In *Proceedings of ArgMAS 2004*, I. Rahwan, P. Moraitis, C. Reed (eds.). Berlin: Springer, pp. 78-94.

Rao, A.S., Georgeff, M. 1991. “Modelling Rational Agents within a BDI-architecture”. In *Principles of Knowledge Representation and Reasoning: Proceedings of KR91*, J. Allen, R. Fikes, E. Sandewall (eds.), pp. 463-484. San Mateo, California: Morgan Kaufmann Publishers.

Belief dynamics, framing effects and decision-making

Natalie Gold

Duke University, goldnk@duke.edu

Not only the beliefs that people have, but also how they manipulate them in the decision making process, can affect the decisions that they make. Whilst, in the-long term, the dynamics of belief change are important, in the short-term, the dynamics of belief usage may be of greater relevance. In this paper, I present a model of the reasoning process developed in Gold and List (2004). The model was motivated by empirical evidence of “framing effects” but, in explaining these, it also shows how framing may more generally enter decision making, so that “frame” change may affect decisions, even when the agent’s belief set remains constant.

There is empirical evidence that changing the representation of a decision problem may affect the choices that people make in it (e.g. Kahneman and Tversky, 1979; Tversky and Kahneman, 1986). This is a framing effect. In particular, choices may depend on the way in which options are described: they are not always description invariant. In a logician’s language, two decision problems may be extensionally equivalent and yet lead to different choices. If we take a descriptive expression from a proposition and substitute a different expression that designates the same object this should, ideally, not affect the truth-value an agent assigns to the proposition. The model uses the framework of predicate calculus to examine exactly which classical conditions of rationality it is whose violation may lead to framing effects. Under one interpretation, it is a model of the process of decision making, so it connects the structure of beliefs with the psychology of decision making.

In the model, an agent may consider several “background” propositions in the run-up to making a decision on a “target” proposition. These background propositions parallel the notion of a reason for choice. I show that, in the model, the agent exhibits a framing effect if and only if two conditions are met. First, different presentations of the decision problem lead the agent to consider the propositions in a different order (the empirical condition). Second, different such "decision paths" lead to different decisions on the target proposition (the logical condition). The logical condition is satisfied if and only if the agent's initial dispositions on the propositions are implicitly inconsistent – which may be caused by violations of deductive closure.

The model has various interpretations. One is that it is a model of the process of reasoning. In this case, the suggestion is that the framing of the decision problem may make particular propositions "available" to the agent and thereby induce the decision-path, in line with the psychological literature on priming, where experimenters' treatments aim to make mental concepts accessible, without subjects being consciously aware that they are being manipulated, and this affects subject's behaviour (Bargh and Chartrand, 2000). It also resonates with the ideas that people act for a single reason (Montgomery, 1983), that only salient information is taken into account in decision making (Slovic, 1972) and that people do not realize that, in another framing of the problem, they would probably have made a different decision (Tversky and Kahneman, 1981). What matters for the decision, in the model, are the particular propositions (reasons) that occur on the decision-path, but not other propositions outside the decision-path even if these seem also relevant from the perspective of an external observer. We might identify those propositions that occur on the agent's decision-path with those beliefs that are in the agent's frame at the time of making a decision.

Hence not only the agent's belief set, but also the frame that she uses, is important in decision making. Some implications and extensions of the model are considered.

REFERENCES

Bargh, J. and Chartrand, T. (2000) 'The Mind in the Middle: A practical guide to priming and automaticity research' in H. Reiss and C. Judd (eds.) Handbook of Research Methods in Social and Personality Psychology New York: Cambridge University Press.

Gold, N. and List, C. (2004) 'Framing as Path-Dependence' Economics and Philosophy 20(2), 253-77.

Kahneman, D. and Tversky, A. (1979) 'Prospect Theory: an Analysis of Decision Under Risk'. Econometrica 47, 263-91.

Montgomery, H. (1983) 'Decision Rules and the Search for a Dominance Structure: Towards a Process Model of Decision Making'. In P.Humphreys, O. Svenson and A. Vari (eds.) Analysing and Aiding Decision Processes, 343-369. Amsterdam: North Holland.

Slovic, P. (1972) 'From Shakespeare to Simon: speculations - and some evidence - about man's ability to process information' ORI Research Monograph 12, (2).

Tversky, A. and Kahneman, D. (1986) 'Rational Choice and the Framing of Decisions' Journal of Business 59, S251-78.

Tversky, A. and Kahneman, D. (1981) 'The Framing of Decisions and the Psychology of Choice'. Science 211, 453-8.

Data, beliefs, and acceptances: Ontology and dynamics of doxastic states

Fabio Paglieri

University of Siena, paglieri@media.unisi.it

The literature on belief dynamics, and especially formal models of them, often failed to acknowledge some relevant aspects of belief formation and change in human cognition, as well as few basic distinctions, drawn in philosophical epistemology, among different mental states related to the subjective assessment of external reality (Alchourrón et al., 1985; Gärdenfors, 1988; Meyer, van der Hoek, 1995; Segerberg, 1999; Pollock, Gillies, 2000; Rott, 2001). The purpose of this contribution is to rectify this misconception, starting from the latter point, i.e. the need for a precise ontology of doxastic states, as a necessary precondition to a proper understanding of belief dynamics. I will focus my attention on two co-related issues: the debate on the definition and properties of *beliefs and acceptances* (van Frassen, 1980; Stalnaker, 1984; Cohen, 1989; Bratman, 1992; Ullmann-Margalit, Margalit, 1992; Engel, 1998; 2000; Tuomela, 2000; Wray, 2001; Tollefsen, 2003), and the distinction between *data and beliefs* (Rescher, 1976; Castelfranchi, 1996; 1997; Tamminga, 2001; Paglieri, 2004). The essential rationale of this preliminary clarification might be summarized as follows: as long as our goal is to consider, compare and assess different models of belief dynamics, we first need to know precisely what we are talking about – that is, what beliefs are supposed to be per se, and what is their place among other doxastic features.

My analysis will be articulated as follows:

- quick review of the distinction between *belief* and *acceptance* as it came to be represented (and debated) in the literature;
- outline of an operational definition of belief and acceptance as different *functions* of doxastic states, namely, truth-functional value and pragmatic value;
- discussion of (i) the import of such definition for *practical reasoning*, and (ii) its place in human *cognitive development* (Robinson, Robinson, 1982; Perner, 1995; Robinson, 2000; 2003);
- introduction and discussion of the distinction between *data* and *beliefs*;
- outline of a cognitive model of belief dynamics as an *emergent effect* of data manipulation (DBR: Data-oriented Belief Revision);
- short discussion on the place of *knowledge* in this framework (if any), with special emphasis on its relations with the concept of belief as understood in cognitive psychology.

REFERENCES

Alchourrón, C., Gärdenfors, P., Makinson, D. (1985). "On the logic of theory change: Partial meet contraction and revision functions". *Journal of Symbolic Logic* 50, pp. 510-530.

Bratman, M. (1992). "Practical reasoning and acceptance in a context". *Mind* 101, pp. 1-15.

Castelfranchi, C. (1996). "Reasons: Belief support and goal dynamics". *Mathware & Soft Computing* 3, pp. 233-247.

Castelfranchi, C. (1997). "Representation and integration of multiple knowledge sources: Issues and questions". In V. Cantoni, V. Di Gesù, A. Setti, D. Tegolo (eds.), *Human & Machine Perception: Information Fusion*. New York: Plenum Press, pp. 235-254.

Cohen, L. J. (1989). "Belief and acceptance". *Mind* 98, pp. 367-389.

Engel, P. (1998). "Believing, holding true, and accepting". *Philosophical Explorations* 1, 140-151.

Engel, P. (ed.) (2000). *Believing and accepting*. Dordrecht: Kluwer.

Gärdenfors, P. (1988). *Knowledge in flux: Modelling the dynamics of epistemic states*. Cambridge, MA: MIT Press.

Meyer, J.-J. C., van der Hoek, W. (1995). *Epistemic logic for AI and computer science*. Cambridge: Cambridge University Press.

Paglieri, F. (2004). "Data-oriented belief revision: Towards a unified theory of epistemic processing". In E. Onaindia, S. Staab (eds.), *STAIRS 2004: Proceedings of the Second Starting AI Researchers' Symposium*. Amsterdam: IOS Press, pp. 179-190.

Perner, J. (1995). "The many faces of belief: Reflections on Fodor's and the child's theory of mind". *Cognition* 57, pp. 241-269.

Pollock, J. L., Gillies, A. S. (2000). "Belief revision and epistemology". *Synthese* 122, pp. 69-92.

Rescher, N. (1976). *Plausible reasoning*. Assen: Van Gorcum.

Robinson, E. J. (2000). "Belief and disbelief: Children's assessments of the reliability of sources of knowledge about the world". In K. P. Roberts, M. Blades (eds.), *Children's source monitoring*. Mahwah, NJ: LEA, pp. 59-83.

Robinson, E. J. (2003). "Six-year-olds' contradictory judgements about knowledge and beliefs". *Trends in Cognitive Science* 7, pp. 235-237.

Robinson, E. J., Robinson, W. P. (1982). "Knowing when you don't know enough: Children's judgements about ambiguous information". *Cognition* 12, pp. 267-280.

Rott, H. (2001). *Change, choice and inference: A study of belief revision and nonmonotonic reasoning*. Oxford University Press: Oxford.

Segerberg, K. (1999). “Two traditions in the logic of belief: Bringing them together”. In H. J. Ohlbach, U. Reyle (eds.), *Logic, language, and reasoning*. Dordrecht: Kluwer, pp. 135-147.

Stalnaker, R. (1984). *Inquiry*. Cambridge, MA: MIT Press.

Tamminga, A. (2001). *Belief dynamics: (Epistemo)logical investigations*. ILLC dissertation series DS-2001-08. Amsterdam: ILLC-UvA.

Tollefsen, D. P. (2003). “Rejecting rejectionism”. *Protosociology* 18-19, pp. 389-405.

Tuomela, R. (2000). “Belief versus acceptance”. *Philosophical Explorations* 2, 122-137.

Ullmann-Margalit, E., Margalit, A. (1992). “Holding true and holding as true”. *Synthese* 92, pp. 167-187.

van Frassen, B. C. (1980). *The scientific image*. New York: Oxford University Press.

Wray, K. B. (2001). “Collective belief and acceptance”. *Synthese* 129, pp. 319-333.

Thursday 17 November 2005, 15:00-19:15, second session:

2.2. Dynamics of knowledge in childhood: Cognitive and developmental perspectives

Chair: Boicho Kokinov (New Bulgarian University)

Belief formation and change in human development

Elizabeth Robinson

University of Warwick, E.J.Robinson@Bham.ac.uk

I shall argue that to understand age related differences in belief formation and change, we need to take into account children’s developing understanding about how minds work. This is particularly true for knowledge gained from other people: humans have the huge advantage over other creatures of being able to gain knowledge from each other in addition to gaining knowledge directly from their own experience of the physical world. However this ability to gain knowledge from others brings with it risks, since other people can deliberately deceive, can unintentionally be in error, and can be misinterpreted. To maximize their chances of believing only what is true, children and adults need to pay attention to characteristics of the speaker as well as to the content of the message.

The evidence suggests that from quite early on, young children are sensitive to cues of reliability or unreliability in others. For example, 3 year-olds do not learn new object names from a speaker who has previously mis-named a familiar object: a speaker’s previous output is taken as a guide to the likely reliability of their future output (e.g. Koenig, Clement, & Harris, 2004). In addition, children this age take into account the relevant information access of a speaker when deciding whether or not to believe what is said: they are inclined to disbelieve a speaker who tells them the colour of an object which the speaker has only felt and not seen (Robinson & Whitcombe 2003; Whitcombe & Robinson, 2000). Even more impressively, 3 year- olds are willing to revise beliefs gained from a speaker who appeared to be reliable at the time, but whose reliability was subsequently called into question (Robinson & Haigh, unpublished). That is, they seem not only to pay attention to the source of their knowledge at the time beliefs are acquired, and also to hold onto that source information at least for a short time afterwards.

Young children achieve all this without yet being able explicitly to reflect on how they know something, and without being able to make explicit judgments of who knows what (e.g. Wimmer, Hogrefe & Perner, 1988). For example, it is not until around 4-5 years that children can predict that to find out the colour of an object they need to see it rather than feel it, or can judge explicitly that someone who has only felt the object does not know what colour it is (O’Neill, Astington & Flavell, 1992; O’Neill & Chong, 2001)

Despite the evidence that 3 year-olds do not show blind trust in whatever they are told, there is some evidence that once 4 to-5 year-olds achieve explicit understanding about how knowledge is gained, they are less suggestible. For example, the literature on eye-witness testimony suggests that children who realise that people can hold false beliefs, and that somebody who was absent when an event happened can be mistaken, may be less inclined to believe an adult's misleading suggestion about an event that the child herself had witnessed (e.g. Welch-Ross, 2000).

However at the age of 4 to 5 years children's understanding about knowledge is still quite rudimentary compared with adults'. For example, when the information available is limited so that we can know possibilities but not certainties, children this age are particularly inclined to over-estimate what they know. They often fail to seek further clarifying information when it is easy to do so, and choose to make a single interpretation rather than hold possibilities in mind (e.g. Beck & Robinson, 2001). This applies both to information from other people (for example an ambiguous utterance which fails to make the intended meaning clear), or information from the physical world (for example a distant object which cannot be identified with confidence). By the age of 7 to 8 years children show more adult-like behaviour in such circumstances. One way of characterizing this development is that children begin to differentiate interpretations of information from the input itself; they realise that a particular input can permit more than one interpretation (e.g. Chandler, Hallett & Sokol, 20020).

Having realised this, they can begin to understand that different people can hold different beliefs about the same input, and begin to understand the active role of the mind in interpreting experiences from the physical world (e.g. Apperly & Robinson, 1998; 2001). That is, instead of treating knowledge and beliefs as simple copies of events in the outside world, they can treat them as active interpretations of events which can be biased by an individual's expectations or prior experiences. For example, if one child knocks over another, they might accept that this could be interpreted as deliberate or accidental depending on the observer's prior experience of the child in question (Pillow & Weed, 1995).

Even this does not mark the end point of development, however. I shall finish by mentioning very briefly research on adults' conceptions of knowledge and belief formation. This focuses on complex inputs about which some people might argue there is no one true interpretation. For example, adults can be exposed to conflicting accounts in the media about complex knowledge concerning scientific, historical or political matters. In such cases, the individual's knowledge is gained entirely from indirect information from others. Some research suggests that for such matters, many adults adopt a relativist stance, arguing that alternative views are equally legitimate and there is no one truth of the matter (e.g. Kuhn, 2000).

REFERENCES

Apperly, I.A., & Robinson, E.J., (1998). Children's mental representation of referential relations. *Cognition, 67*, 287-309.

Apperley, I.A., & Robinson, E.J. (2001). Children's difficulties handling dual identity. *Journal of Experimental Child Psychology, 78*, 374-397.

Beck, S. R. & Robinson, E.J. (2001). Children's ability to make tentative interpretations of ambiguous messages. *Journal of Experimental Child Psychology, 79*, 95-114.

Chandler, M.J., Hallet, D., & Sokol, B.W. (2002). Competing claims about competing knowledge claims. In B. Hofer and P. Pintrich (Eds.) *Personal epistemology: The psychology of beliefs about knowledge and thinking* (pp. 145-167), Mahwah, NJ, Lawrence Erlbaum.

Koenig, M. A., Clement, F. & Harris, P. L. (2004). Trust in testimony: Children's use of true and false statements. *Psychological Science, 15,* 694-698.

Kuhn, D. (2000). Theory of mind, metacognition, and reasoning: A life-span perspective. In P. Mitchell & K. Riggs (Eds.), *Children's reasoning and the mind* (pp. 301-326). Hove. England. Psychology Press

Naito, M. (2003). The relationship between theory of mind and episodic memory: evidence for the development of autonoetic consciousness. *Journal of Experimental Child Psychology, 85*, 312-336.

O'Neill, D. K., Astington, J. W. & Flavell, J. H. (1992). Young children's understanding of the role that sensory experience plays in knowledge acquisition. *Child Development, 63*, 474-490.

O'Neill, D K. & Chong, C. F. (2001). Pre-school children's difficulty understanding the types of information obtained through the five senses. *Child Development, 72*, 3, 803-815.

Pillow, B.H. & Weed, S.T. (1995). Children′s understanding of biased interpretation: Generality and limitations. *British Journal of Developmental Psychology, 13,* 347-366.

Robinson, E. J. & Whitcombe, E. L. (2003). Children's suggestibility in relation to their understanding about sources of knowledge. *Child Development, 74*, 48-62.

Welch-Ross, M.K. (2000) Pre-schoolers' understanding of the mind: Implications for suggestibility. *Cognitive Development, 15*, 101-131.

Whitcombe, E. L. & Robinson, E. J. (2000). Children's decisions about what to believe and the ability to report the source of their beliefs. *Cognitive Development, 15*, 329-346.

Wimmer, H., Hogrefe, G. J. & Perner, J. (1988). Children's understanding of information access as a source of knowledge. *Child Development, 59*, 386-396.

Knowledge acquisition and conceptual change in childhood

Stella Vosniadou

University of Athens, svosniad@phs.uoa.gr

For the last 10 years we have been involved in research investigating the problem of knowledge acquisition and conceptual change in the areas of the physical sciences and mathematics. Our studies are experimental and use mainly two kinds of methodologies – cross-sectional developmental experiments and instructional interventions with measures of pre and post learning. There have also been some attempts to create computational models of conceptual change in the process of learning science (see Kayser &Vosniadou, 2000). The picture of the long term dynamics of knowledge during learning and development that emerges from these studies and that we will argue for is the following:

a) The knowledge acquisition process starts soon after birth and develops in an orderly fashion along certain high level domains (i.e., physics, psychology, number)
b) By the end of the preschool years children have formed weak framework theories that strongly constrain further knowledge acquisition processes
c) Learning science and mathematics requires conceptual re-organisation that can roughly be described in terms of theory change
d) The mechanisms of knowledge acquisition mostly used by young learners are additive and aim at the enrichment of prior knowledge structures.
e) The use of such mechanisms can explain the creation of misconceptions in science and mathematics. According to this view, misconceptions are synthetic models formed as new information coming from the culture is added on to existing but incompatible knowledge structures, creating hybrid models
f) Some form of metaconceptual awareness is necessary for the development of more sophisticated knowledge acquisition mechanisms that depend on hypothesis testing and the conscious exploitation of analogy (although analogical reasoning can operate from much earlier on).

REFERENCES

Kayser, D., & Vosniadou, S. (Eds.) (2000). Modelling changes in understanding: Case studies in physical reasoning, Elsevier

Cognitive development: the balance scale task

Han van der Maas

University of Amsterdam, h.l.j.vandermaas@uva.nl

The balance scale task, a task for proportional reasoning, plays an important role in the study of cognitive development. In this task children have to predict the movement of a balance scale with different configurations of weights located somewhere on four equidistant pegs on each side of the fulcrum. It has been shown that children use rules or strategies in solving his task. Young children, for instance, ignore the distances and only take weight into account. Somewhat older children, use distance information only when the number of weights on each side is equal. Older children use more advanced strategies, including the correct one based on the comparison of the torques. Alternatively, they may compare the sums, a popular strategy that succeeds on the large majority of balance scale items. An advantage of this task is that it can be used with 4 year-old children but also with much older subjects. Many adults still fail this task.

The balance scale task has been used as an important benchmark task for computational models (both symbolic and connectionist) for cognitive development. Well-known are the connectionist models of the balance scale task, the PDP model of McClelland and the cascade correlation model of Shultz. An important discussion concern the question whether children really use rules or that their seemingly rule like behavior can well be mimicked by connectionist networks. In this talk I will present a number of new empirical results concerning a) applications of new statistical techniques to rule assessment (latent class analysis) b) phase transitions in development (hysteresis) c) advanced analysis of response times d) longitudinal data about rule change. Based on these new data I will argue that children, in spite of some irregularities, indeed use cognitive rules. Current connectionist networks fail to explain this, although new connectionist models may do a better job. A recently proposed Act-R model does explain most empirical phenomena. This model is based on a very general strategy: “looking for differences”. Our general conclusion is that cognitive development can be characterized by a domain specific sequence of increasingly complex rules or strategies for solving problems.

Young children’s intuitions about posterior probability

Vittorio Girotto[1], Michel Gonzalez[2]

(1) University IUAV of Venice, girotto@mercurio.univ.trieste.it
(2) CNRS / University of Provence

Do young children possess basic intuitions of posterior probability? Do they update their judgments in the light of new evidence? We hypothesized that they can do so extensionally, by considering and counting the various ways in which an event may or may not occur. The results reported in this series of studies showed that from the age of five, children’s choices (Study 1) and judgments (Study 2) about random outcomes are correctly affected by posterior information. From the same age, children correctly revise their choices in situations in which they have to reason about a single, not random outcome (Study 3). The finding that young children have some correct intuitions of posterior probability supports the theory of naïve extensional reasoning, and contravenes some pessimistic views of naive probabilistic reasoning.

On the stability of instruction and experience based beliefs

Kristien Dieussaert, Deborah Vansteenwegen, An Van Assche

Catholic University of Leuven, kristien.dieussaert@psy.kuleuven.be

Introduction

Research on belief revision has only very recently become a topic of interest within human reasoning research. For a review of the theories and recent data in a special issue on reasoning from inconsistency, we refer to Dieussaert and Schaeken (2005).

Generally, participants are given a conditional statement (if p, then q; e.g., if that bacteria is present in your blood, then you have the Okro disease) and a categorical statement (p; e.g., the bacteria is present), and are asked to deduce the conclusion, or are given the conclusion (q; e.g., you have the Okro disease). Next, new information that contradicts the conclusion is given (not-q; e.g., you do not have the Okro disease) and participants are asked to revise one of the former statements in order to regain a consistent belief set.

One of the subdomains of psychology in which particularly interesting research related to belief revision is conducted, is that of contingency learning. For a review, see De Houwer and Beckers (2002). In most human contingency learning experiments, participants receive information about a number of situations in which certain Cues (C) and Outcomes (O) are either present or absent, and they are asked to judge the extent to which the presence of a C is related to the presence of O. On the basis of this information participants will be able to formulate a rule about the C-O relation.

In reasoning research, the learning part is restricted to the presentation of the established relationships in the form of conditional statements (If C, then O) or universal quantifiers (All C's are/have/.. O's). The similarity between the C-O relations and conditional statements (If C, then O) is obvious. Knowledge about the principles and circumstances under which C-O relations are acquired and extinguished can lead to fruitful insights in how belief states are constructed and revised and vice versa.

As may be clear now, both research areas use a different experimental paradigm to induce a belief in a C-O relation. These operationalisations reflect a different view on how beliefs are constructed: through instruction (if C, then O) or through experience (several C-O trials). We consider both forms of belief construction important since people construct their beliefs in various ways, depending on the situation. Some beliefs are constructed through communication (e.g., If you run out of brake oil, your brake will not work) while others are constructed through experience (e.g., If you eat, your hunger stops).

The main goal of the studies that will be presented is to find out whether the methodology of belief construction affects the belief strength and whether it affects the belief revision process. In other words, does a theory driven or a data driven belief construction give rise to a more entrenched belief state? As a case study, we focused on a rather recent discovery in human contingency learning, viz. the phenomenon of 'renewal' (e.g., Garcia-Gutierrez & Rosas, 2003): the return of an extinguished C-O relation due to context change. Translated in terms of the reasoning process, renewal refers to a (renewed) expression of someone's belief in the conditional sentence (if C, then O), despite the presence of contradictive information (co-occurrence of C and not-O).

Experimental evidence

Belief strength.

Belief was induced in three ways: Instruction (I), Experience (E), and Instruction+Experience (IE). If we compare the three groups after one trial, we observe the lowest belief strength in the Egroup (M = 5.36 on a seven point scale) and the highest belief strength in the IE group (M = 6.34; $p < .0001$). The E group scores also lower on belief strength than the I group (M = 5.94; $p < .06$). However, already after two experience trials the difference in belief strength between the three groups disappears completely (M E group = 6.09).

Context change.

When the context is changed, so that it differs from the context in which the belief was acquired, participants tend not to give up their former beliefs, although they express some doubt on how to react in such a situation. This results in a drop in belief score (M = 4.95). The context change has a similar effect for the three belief induction groups.

Belief contradiction.

If the belief contradiction is induced through instruction, its effect is immediate: the belief score drops (M = 1.93). If the belief contradiction is induced through experience, the effect is more gradually (M =2.96 after one negative trial, M = 2.31 after two negative trials, …). In case no context change is paired with the belief contradiction, the pattern remains the same: a major drop with contradiction through instruction (M = 2.63) versus a gradual decrease with contradiction through experience (M = 4.35 after one negative trial, M = 2.90 after two negative trials, …).

Renewal.

Renewal is found in all groups: when the original, learning context pops up, the belief strength increases again (M = 4.42). However, the renewal is only complete when belief acquisition and contradiction happened through experience (M = 5.07). Particulary interesting is that the renewal is least when contradiction was presented through instruction (M = 3.15), more when it was presented through instruction combined with experience (M = 4.07), and most when it was presented through experience (M = 4.77).

Discussion

Although the first results indicate that belief strength differs depending on the acquisition procedure, this effect fades away if learning through more than a single experience is allowed. Instruction and experience do have a different effect on the stability of beliefs however, in that there are many indications that instructions in the form of a conditional rule (if C, then O) are more prone to generalisation over various contexts than subsequent experiences of C-O trials.

REFERENCES

De Houwer, J. & Beckers, T. (2002). A review of recent developments in research and theories on human contingency learning. *Quarterly Journal of Experimental Psychology, 55B* (4), 289-310.

Dieussaert, K., & Schaeken, W. (2005). Reasoning from inconsistency: a field exploration. *Psychologica Belgica, 45* (1), 1-18.

Garcia-Gutierrez, A. & Rosas, J. M. (2003). Empirical and theoretical implications of additivity between reinstatement and renewal after interference in causal learning. *Behavioural Processes, 63* (1): 21-31.

Friday 18 November 2005, 9:00-13:00, third session:

2.3. Evolutionary and socio-cognitive dynamics of knowledge

Chair: Cristiano Castelfranchi (ISTC-CNR Roma / University of Siena)

The evolution of a theory of mind, common knowledge and cooperation

Peter Gärdenfors

Lund University, peter.gardenfors@lucs.lu.se

In my recent book *How Homo became Sapiens* (Oxford University Press 2003), I present a hierarchy of levels of ”theory of mind” in the evolution and ontogeny of thinking. The highest level is to *understand the beliefs of others*, which is only attained by humans and ontogenetically only at about four years of age. This mental capacity is necessary for achieving *common knowledge* of the kind that one finds in human social structures such as language, monetary systems and all kinds of conventions.

Common knowledge is a social phenomenon that cannot be located in the heads of single individuals. Nevertheless, common knowledge has *causal powers* that emerge from the interactions of the individuals and their beliefs. The analogy with Wiener's virtual governor is often applied to elucidate what kind of causation is involved.

Game theory has been designed to analyse the dynamics of beliefs in cooperative and non-cooperative. A central concept in the theory is that of an *equilibrium* strategy, that can be seen as an emergent cause in a system based on individual beliefs and desires. However, the role of a theory of mind and common knowledge is, in general, neglected in this theory (one exception is Schelling's coordination games).

In contrast, I believe that common knowledge should be much more exploited in game theory. For example, in the classical theory, a strict partitioning between cooperative and non-cooperative games is made. In real life, the two extremes of co-operation and non-cooperation are rarely attained. In most cases a player has only partial information about the choices and potential behaviour of his opponents, either as a result of memories from earlier, similar situations (for instance, in iterated games) or as a consequence of other kinds of expectations, most typically based on a theory of mind.

For example, in a prisoners' dilemma (PD) game the players have two options — to cooperate or to defect. If the PD is seen as a purely cooperative game, traditional game theory prescribes the cooperative strategy as the only rational one for all players. In contrast, in a purely non-cooperative PD, the theory claims that defecting (the non-cooperative strategy) is the only rational strategy.

In real situations where the game is described as a non-cooperative one, human subjects (and animals) often choose the cooperative strategy, in contrast to what is recommended by game theory. The reason for this seeming irrationality is that a PD type situation is seldom treated as a strictly non-cooperative game by the subjects. Even if a subject does not have any real information about her opponents' choices in the situation, she has expectations about their behaviour. For example, she may count on that they reason in the same way as she does herself, or that they would, like herself, feel ashamed if they chose the defecting strategy.

Such "theory of mind" expectations function as information about the choices of the others that effectively make the game situation partly cooperative. In such a situation, the rational move to make may very well be to cooperate. Since it is hard to imagine a game situation where a human player has no expectations whatsoever about the opponents, it is questionable whether the pure non-co-operative situation prescribed in game theory can ever be attained.

Another factor that influences the choice situations in PD type games is that among social animal species, and humans in particular, the possibility of *sanctions* from the rest of the group may drastically change the game situation. Even if you temporarily gain by defecting in a (non-iterated) PD situation, the risk of being punished by the peers in the group for such a non-cooperative (egoistic) behaviour should be taken into account when calculating the utilities of the available strategies. If the punishment is severe and the risk of receiving it high enough, the payoffs of the game will change in such a way that it no longer is a PD, but a game where the only rational strategy is to cooperate. Consequently, including expectations about (long term) sanctions is a way of changing the rational equilibrium of a PD type game into a game with only a co-operative equilibrium.

The functioning of sanctions depends on another uniquely human form of cognition: *anticipatory thinking*. A sanction promises a punishment in the future for a socially non-desired action that is performed now. An agent who has no concept of the future cannot be influenced by a sanction.

Another aspect of thinking about the future is anticipatory planning. Humans, but no other animals, can engage in planning for goals that do not exist at present. An additional effect on PD situation is that the presence of shared knowledge (or beliefs) about a future goal will by change a situation, which would be a PD without the presence of such knowledge, into a game where the cooperative strategy is the equilibrium solution. For example, if somebody communicates the idea that we should cooperate in digging a communal well, then such a well, by being deeper, would yield much more water than all the individual wells taken together. Once such cooperation is established, the PD situation may disappear, since everybody will benefit more from achieving the common goal. In game theoretical terms, digging a communal well will be a new equilibrium strategy. This example shows how the capacity of having common beliefs about future goals can strongly enhance the value of cooperative strategies within the group. The upshot is that strategies based on future goals may introduce new equilibria that are beneficial for all participants.

Evolution of conceptual maps as a result of learning

Maurice Grinberg

New Bulgarian University, mgrinberg@nbu.bg

A methodology for the evaluation of conceptual change based on mental maps is proposed and analyzed. The main idea of this methodology consists in the application of mental maps to assess the conceptual structure of students in different moments in time, and analyze the differences between them. The evolution of conceptual structures can be traced in time by comparing subsequent mental maps or by comparing to a reference map obtained with the help of experts or by means of latent semantic analysis (LSA) or a similar analysis. The mental maps are established on the basis of a free classification task followed by a hierarchical cluster analysis and multidimensional scaling. This methodology allows exploring the various groups of concepts that arise, the between group and within group distances and their evolution in time both qualitatively and quantitatively. The information obtained can be related to relatively important (discrete) changes in the mental maps (e.g. changes of concept group members) or to more continuous ones like changes in the distances between the concepts within a group.

As far as the processes of conceptual change are especially important in learning, the methodology was applied to the study of conceptual change in the domain of computer science in high-school and in first year university psychology. The conceptual maps for concepts from the domain of computer science of school students have been obtained on the basis of similarity judgments. First, a representative list of terms was generated on the basis of the students' curriculum in computer science. Then the participants were asked to group the terms, following the requirements of a free classification task, into as many categories as they wanted, keeping in mind their similarity in meaning. The free classification task was carried out with the participation of four consecutive classes of students who have studied computer science in the same school. Based on the grouping task similarity matrices for the terms included were obtained for each class of students. Additional similarity matrices for the same list of terms were obtained by performing the same experiment with advanced students from a school with a specialization in computers. They latter could be considered as experts relative to the terms used in the experiment which were taken from the general computer literacy domain. The similarity matrices obtained were analyzed by different statistical methods like hierarchical cluster analysis and multidimensional scaling producing mental maps. This same procedure was applied to study the change in knowledge in general psychology between first and second year university students in psychology.

The mental maps thus obtained are compared among them selves and with a reference map generated by LSA on the basis of text books. As expected, these analyses exhibit differences among the similarity matrices which can be interpreted as an improvement of the conceptual structures as compared to a target level established for instance by LSA.

The results presented in the paper show that the proposed method of analysis of the time evolution of knowledge in a single domain can give important information about the dynamics of the process of concept acquisition and conceptual structure change.

Multi-player belief revision and information value in games

Antoine Billot[1], Jean-Christophe Vergnaud[2], Bernard Walliser[3]

(1) PSE-ENPC, CORE
(2) CNRS Eureqa
(3) PSE-ENPC, EHESS Paris, walliser@enpc.fr

A multi-player belief structure expresses the (crossed) beliefs of a set of players about the physical world in standard epistemic logics. It is formalized in a syntactical framework (propositions and belief operators) as well as in a semantical one (possible worlds and accessibility relations). Together with an initial belief structure, one considers a message formed of two elements. The 'content' of the message makes precise what each player learns among a set of possibilities. Such a message may be material (about the physical world) or epistemic (about other agents' beliefs). It is expressed in the very terms of the initial belief. The 'status' of the message describes to what player the message is sent and what the other players know about the message diffusion (at all levels of crossed beliefs). Among other instances, a message may be public (each player receives a message and this is common knowledge), private (one player only gets the message, but this is common knowledge) or secret (one player only gets the message, and this is not known by the others). The status of the message is formalized by an auxiliary belief structure, which is always expressed in a syntactic or a semantic framework. The belief revision process turns the initial belief structure and the message into a final belief structure. A unique revision axiom is given in syntax and the corresponding revision rule is deduced in semantics through a representation theorem.
Moreover, a bisimilarity relation expresses that two belief structures are semantically equivalent when they lead to the same syntactical structure. Moreover, three types of accuracy relations are defined between two belief structures, formalizing the intuitive fact that some player knows 'more' in the first structure than in the second one. These relations are stated on beliefs expressed in a syntactical framework (by inclusion of proposition sets) and transposed into a semantical one (by inclusion of accessibility domains). Coming back to the belief revision process, it can be shown that, given an initial belief structure, if the status of a message is more accurate (in some sense) than the status of another, the corresponding final belief is also more accurate (in the same sense).
Furthermore, a game structure is introduced in which the players'uncertainty is expressed by a probabilized belief structure about some physical states of nature. A 'semantic equilibrium' is defined in a Nash fashion. The information value brought by a message to some player is defined as the (average) difference in utility he gets at equilibrium before receiving the message and after receiving the message. Under some technical conditions, the information value is proved to be positive in three cases : (i) for the receiver of a secret message, (ii) for the receiver of a private message in a zero-sum game, (iii) for all players receiving a public message in a pure coordination game.

Knowledge in Economics

Brian J. Loasby

University of Stirling, b.j.loasby@stir.ac.uk

In the second half of the twentieth century theories of choice in economics were oriented towards deriving equilibria of optimising agents. In the first part of this period the emphasis was on general equilibrium, because this was thought appropriate for analysing the allocation of resources within an economic system, but more recently the focus has switched to various kinds of interaction between agents, to be modelled as Nash equilibria. In neither class of models is knowledge typically treated as problematic: information may be asymmetrically distributed (indeed this is a standard assumption for some classes of problem); but it is normally assumed that everyone knows the implications of this asymmetry. This conforms with the current tendency to conceive equilibria in terms of internal consistency. Herbert Simon's criticisms have either been rejected or interpreted as minor qualifications – for example by replacing full information by information which is optimally selected from a known possibility set.

There is, however, another tradition in economics which takes economics more seriously, and which, not surprisingly, leads to a substantially different view of the working of economic systems. For the purposes of this interdisciplinary workshop the most efficient use of my time is probably not to consider the treatment of knowledge in economic systems but to outline and comment on the ideas about the working of the human mind developed – quite independently – by three famous economists early in their careers – indeed before they took up their study of economics. All developed evolutionary theories which respected the scarcity of human cognitive powers and relied on the formation and modification of selective connections within the mind. The three were Adam Smith, Alfred Marshall, and Friedrich Hayek. In order to develop my argument I shall take them in the reverse of chronological order.

Hayek's enquiry was motivated by the incommensurabilities between the sensory and physical orders of a class of phenomena. He argued that the human brain develops classification systems through the increasingly regular transmission of particular impulses within the central nervous system, and that it is therefore physically possible to develop alternative neural systems which classify a set of phenomen in different ways. His exposition has been recognised as a substantial contribution to theoretical psychology.

Marshall responded to the work of Darwin, Spencer, Babbage and the less well-known Alexander Bain, who produced the first major treatment of physiological psychology, by producing a mental model of a 'machine' with a mechanical brain which built up connections between 'ideas of impressions' and 'ideas of action' in the course of interactions with the environment. This happened in two circuits, first by simple trial and error and second by a more cognitively-expensive process which included conjecture and (fallible) pre-selection: at each level successful patterns became encoded as routines, freeing cognitive resources for new conjectures. Marshall's model may be applied to Hayek's system to explain why the sensory order comes first, and is not displaced for ordinary human activity by the physical order.

Smith was influenced by Hume's criticism of induction to develop an emotional-cognitive account of the development of science through the invention of 'connecting principles' which appealed to the imagination rather than to sensory perception, thus supplying, almost two centuries in advance, an explanation for the different classification systems of the sensory and physical orders which prompted Hayek's theory, as well as explaining why these differences should provide an incentive for his enquiry.

Smith went on to show how the progressive differentiation of the sciences accelerated the growth of knowledge by generating increasingly distinctive categories of problems, each with their own criteria for satisfactory explanations, and exploiting the ability of different people to develop quite distinctive structures of knowledge (contrary to the assumptions of the economic theories mentioned in the opening paragraph). This sequence was then transferred to economics to explain why the division of labour was the key to the growth of productive knowledge. The brain and the economy both make extensive use of domain-limited systems of connection – as exemplified notably in Marshall's account of industrial organisation.

REFERENCES

Hayek, Friedrich A. (1952) *The Sensory Order*. Chicago: University of Chicago Press.

Marshall, Alfred (1994) 'Ye Machine'. *Research in the History of Economic Thought and Methodology, Archival Supplement 4*, pp. 116-32. Greenwich CT: JAI Press.

Raffaelli, Tiziano (2003) *Marshall's Evolutionary Economics*. London and New York: Routledge.

Smith, Adam (1980 [1795]) 'The principles which lead and direct philosophical enquiries: illustrated by the history of astronomy', in *Essays on Philosophical Subjects*, ed. W. P. D. Wightman. Oxford: Oxford University Press, pp. 33-105.

Word of Mouth: The Added Value of Beliefs' Dynamics

Paolo Turrini, Mario Paolucci, Rosaria Conte

ISTC-CNR Roma, paolo.turrini@istc.cnr.it, mario.paolucci@istc.cnr.it, rosaria.conte@istc.cnr.it

Knowledge diffusion in a society of intelligent autonomous agents can give rise to emergent and complex inter-agent properties.

The topic of the present talk is social reputation, considered as a fundamental mechanism of social intelligence that allows for the diffusion and evolution of socially desirable behaviours, like altruism, cooperation and norm abiding.

Gossip (the process of transmission of reputation) is seen as a vehicle of strategic knowledge, relevant for partner selection and cheaters detection. The presence of autonomous agents, endowed with filters for both belief and goal adoption and generation, seems intuitively enough to conclude that reputation cannot merely circulate by means of imitation, since agents can intentionally spread fake reputation, and refuse to adopt an evaluation shared by the majority. Actually, in order to adopt an evaluation agents undergo a complex mental process, which can be captured only by considering beliefs and their dynamic.

Several questions need to be answered in order to deal with reputation among intentional agents: Why should agents spread reputation? What are the cognitive ingredients necessary for its transmission?

We claim that cognitive ingredients are necessary to deal with reputation, which is a complex, multifaceted object, resulting from a process of social transmission, consisting of both a factual property and a mental state. In a multidirectional process of emergence, it is both a cause and an effect of social behaviour and, beforehand, of the mental states and processes governing it.

To test these claims, we are carrying on a cross-methodological research. Hypotheses about the role of reputation have been tested in both cooperative settings, by means of multi-agent-based simulative experiments, and in competitive settings, by means of natural experiments in a virtual market. Furthermore, the role of reputation has then been formally analyzed and its added value demonstrated by means of logical argumentation. The previously presented model of reputation has been fed into a logical apparatus, aimed at describing the concepts underlying social reputation theory and at deriving the cognitive ingredients (Beliefs, Desires, Intentions) involved in reputation representation and spreading.

The rest of the talk is organised as follows:

- First the current views of the role of reputation in agent societies are analyzed and criticised, next a new view is proposed which will be presented and developed later on.
- Afterwards, a sociocognitive model of gossip as reputation transmission will be presented and discussed, based upon a fundamental distinction between image (evaluation) and reputation (metaevaluation), both in their social and individual aspects.
- Thereafter, the uses of reputation in cooperation and competition and the reasons for its wide spreading will be shown. Findings from simulative studies about the role of reputation in norm-compliance - a special type of socially desirable behaviour - will be shown to be consistent with the model provided before and to justify the emphasis laid on reputation transmission. These findings will be integrated with relevant results from natural settings.
- Finally a logical analysis of the difference between image and reputation will be provided, in order account for their respective transmissibility. Logical analysis will be aimed at capturing the dynamics of their transmission process both in their cognitive and social

aspects. Possibilities for a link between formalization and computation in evaluation spreading will be at last explored.

REFERENCES

Cristiano Castelfranchi and Isabella Poggi. *Bugie, finzioni e sotterfugi*. Carocci, Firenze, 1998.

Rosaria Conte and Cristiano Castelfranchi. *Cognitive and Social Action*. UCL Press, London, 1995.

Rosaria Conte. Memes through (social) minds. In R. Auger, editor, *Darwinizing Culture: the Status of Memetics as a Science*. Oxford University Press, 2000.

Rosaria Conte and Mario Paolucci. *Reputation in Artificial Societies: Social Beliefs for Social Order*. Kluwer, Dordrecht, 2002.

Rosaria Conte and Mario Paolucci. *Responsibility for societies of agents*. Journal of Artificial Societies and Social Simulation, 7-4, 2004.

R. Dawkins. *The Selfish Gene*. Oxford University Press, 1976.

Robin Dunbar. *Grooming, Gossip, and the Evolution of Language*. Harvard University Press, London, 1998.

Rino Falcone and Cristiano Castelfranchi. *Social trust: A cognitive approach*. In Cristiano Castelfranchi and Yao-Hua Tan, editors, Trust and Deception in Virtual Societibes, Kluwer Academic Publishers, pp 55-90, 2001.

Henry Hexmoor, Cristiano Castelfranchi, and Rino Falcone. *Agent Autonomy*. Kluwer Academic Publishers, Norwell, 2003.

Mario Paolucci. *Reputation as a complex cognitive artefact: theory, simulations, experiments*. PhD Thesis, 2005.

Wiebe van der Hoek and Michael Wooldridge. *Towards a logic of rational agency*. Logic Journal of the IGPL, 11 (2):133–157, 2003.

Friday 18 November 2005, 15:00-19:15, fourth session:

2.4. Formal models of knowledge dynamics: Comparison with social and cognitive theories

Chair: Peter Gärdenfors (Lund University)

Cognition as interaction*

Johan van Benthem

University of Amsterdam / Stanford University, johan@science.uva.nl

*An extended version of this abstract is available on line at http://www.media.unisi.it/cirg/udk/abs_esf_vanbenthem.pdf

Many cognitive activities are irreducibly social, involving interaction between several different agents. We look at some examples of this in linguistic communication and games, and show how logical methods provide exact models for the relevant information flow and world change. Finally, we discuss possible connections in this arena between logico-computational approaches and experimental cognitive science.

When King Pyrrhus of Epirus, one of the foremost well-educated generals of his age, had crossed over to Italy for his famous expedition, the first reconnaissance of a Roman camp near Tarentum dramatically changed his earlier perception of his enemies (Plutarch, "Pyrrhus", Penguin Classics, Harmondsworth, 1973):

> Their discipline, the arrangement of their watches, their orderly movements, and the planning of their camp all impressed and astonished him – and he remarked to the friend nearest him: “These may be barbarians; but there is nothing barbarous about their discipline”.

It is intelligent social life which often shows truly human cognitive abilities at their best and most admirable. But textbook chapters in cognitive science mostly emphasize the apparatus that is used by single agents: reasoning, perception, memory, or learning. And this emphasis becomes even stronger under the influence of neuroscience, as the only *obvious* thing that can be studied in a hard scientific manner are the brain processes inside individual bodies. Protagoras famously said that "Man is the measure of all things", and many neuroscientists would even say that it's just her brain. By contrast, this contribution makes a plea for the irreducibly social side of cognition, as evidenced in the ways in which people communicate and interact. Even in physics, many bodies in interaction can form one new object, such as a solar system. This is true all the more when we have a meeting of many minds!

Perhaps the simplest and yet most striking example of interactive cognitive behaviour is *language use in conversation*. This will be the key example here, as we will discuss a variety of formal approaches to the nature of interaction in language processes: among others, dynamic epistemic logic, information update, belief revision, game-theoretical models, and dynamical systems. Experience in these areas has shown two things. First, there is enough substance to create exact theories – but also, such theories need to take their cues from quite diverse disciplines, such as linguistics, philosophy, logic, computer science, economics, and cognitive psychology. This talk aims to present and discuss some concrete examples of this confluence, all from a logician's perspective.

Analyzing communication in a logic of grounding, belief and intention

Andreas Herzig, Benoit Gaudou, Dominique Longin

IRIT-CNRS Toulouse, Andreas.Herzig@irit.fr

There are two ways to analyze communication: the first is through its structure, and the second is through the participants' mental states. The former route is taken by conventional approaches such as Conte and Castelfranchi's and Walton and Krabbe's, and by social approaches such as Singh's and Colombetti's. These approaches focus on what a third party would perceive without referring to participants' mental states. They thus avoid strong hypotheses on the agents' mental states such as sincerity and cooperation.

We here propose a synthesis of the mental and the structural approach that is based on the notion of grounding. We say that a proposition is grounded if it is expressed and established during a conversation. We define a multi-modal logic that combines the logic of grounding with Cohen and Levesque's logic of belief and intention. Within this logic we formalize Walton and Krabbe's PPD0 persuasion dialogues by characterizing the corresponding speech act types in terms of grounded beliefs and intentions. Our characterization induces a protocol governing the conversational moves.

Shifting priorities: Simple representations for twenty-four iterated theory change operators

Hans Rott

University of Regensburg, hans.rott@psk.uni-regensburg.de

Prioritized bases, i.e., weakly ordered set of sentences, have been used for representing an agent's `basic' or `explicit' beliefs, and alternatively for compactly encoding an agent's belief state (without the claim that the elements of a base are in any sense basic). This paper focuses on the second interpretation of prioritized bases. I explain how the shifting of priorities in such bases can be used for a simple, constructive and intuitive way of representing a large variety of methods for the change of belief states -- methods that have usually been characterized semantically by a system-of-spheres modelling. Among the methods represented are external, radical, conservative an moderate revision, revision by comparison in its raising and lowering variants, as well as various constructions for belief expansion and contraction.

REFERENCES

J. Cantwell: 1997, `On the logic of small changes in hypertheories', Theoria 63, 54-89.

E. Fermé and H. Rott: 2004, `Revision by comparison', Artificial Intelligence 157, 5-47.

A.C. Nayak: 1994, `Iterated belief change based on epistemic entrenchment', Erkenntnis 41, 353-390.

H. Rott: 1991, `A Non-monotonic Conditional Logic for Belief Revision I', in A. Fuhrmann and M. Morreau (eds.), The Logic of Theory Change, LNCS 465, Berlin: Springer, pp. 135--81.

H. Rott: 2003, `Coherence and conservatism in the dynamics of belief II: Iterated belief change without dispositional coherence', Journal of Logic and Computation 13, 111-145.

M.-A. Williams: 1994, `On the logic of theory base change', in C. MacNish, D. Pearce and L.M. Pereira (eds.), Logics in Artificial Intelligence, LNCS 838, Springer, Berlin, pp. 86-105.

Dynamic doxastic logic: Why, whether, how

Krister Segerberg[1]*, Hannes Leitgeb*[2]

(1) Uppsala University, krister.segerberg@filosofi.uu.se
(2) University of Bristol, Hannes.Leitgeb@sbg.ac.at

By combining elements of doxastic logic with the formal theory of belief revision, *dynamic doxastic logic* aims at a unified logical account of rational belief change. In our talk we will scrutinise its underlying research programme, we will question its purpose and viability, and we will survey the ways in which it has been and can be carried out.

The first part of the talk concentrates on whether an object language representation of the belief revision operator in terms of modal operators with a possible worlds representation should be given at all, and, if so, whether non-material conditionals could be the adequate means of doing so. In related areas, such as non-monotonic reasoning and probabilistic update, consequence relations or numerical assignments are often only expressed meta-linguistically. The well-known impossibility results concerning belief revision for languages with conditionals seem to indicate that syntactic items which involve the belief revision operator either do not express propositional beliefs or are not subject to revision in the same sense as factual beliefs are. On the other hand, David Lewis' spheres semantics for counterfactual conditionals is formally very close to the spheres models for standard systems of belief revision. We present some new results which ought to illuminate the formal and philosophical merits or shortcomings of representing "$B \in K^*A$" by conditionals on the object language level. In particular, we show that Arrow's theorem in the theory of social choice can be understood as a restricting result of a similar kind as Gärdenfors' impossibility result and we deal with a new manner of expressing belief revision operators by conditionals with doxastic antecedents and consequents.

The second part of the talk is devoted to the treatment of belief revision operators as the dynamic versions of unary modal operators in doxastic or epistemic logic. In addition to the fact that standard proof-theoretic and semantic methods of modal logic can be applied successfully in the logical analysis of belief change, the generalized point of view which the possible worlds model of modal operators allows can be of great use for the further development of the theory of belief revision. E.g., a separation of the revision axioms into those which are included in any "normal" system and those which characterize particular properties of accessibility relations would be highly attractive. We discuss a new semantical approach in which belief states are analyzed as so-called hypertheories, we indicate how different constraints on preference relations for "fall back" positions can be encoded in this semantical approach, we deal with the logical axioms and rules that

correspond to these constraints, and we show that both belief revision and belief update can be subsumed under this logical framework.
At the end of our talk we review what has been achieved and draw some general conclusions on the structure of belief states and the properties of their dispositional and propositional components.

Mixed moods and unmixable modalities. Modeling beliefs and ntentions

Frank Veltman

University of Amsterdam, veltman@illc.uva.nl

In my talk[1] I will outline a semantics for epistemic and deontic modalities which sheds some light on two problems that so far have been neglected by most of us.

Problem 1. Mixed moods

Most logicians would say that declarative sentences have a truth value, and that imperatives do not. A declarative sentence denotes a proposition, an imperative denotes something else. (There is no agreement about what exactly the denotation of an imperative would be.)
However, if declaratives and imperatives denote different kind of objects then what is the denotation of sentences like 'Stop or I'll shoot' in which these different moods are put together? To deal with sentences like this we need a uniform notion of meaning on which we can base a notion of logical validity that is applicable to declaratives and imperatives, and to sentences in which these moods are combined.
I will argue that the framework of update semantics provides the notions required. In update semantics the meaning of a sentence is equated with the impact it has on the cognitive state of an addressee. A conclusion ψ follows from a sequence of premises $\phi_1, \dots, \phi_n$ if the conclusion ψ has no further impact on the cognitive state of anyone who has learnt the premises $\phi_1, \dots, \phi_n$; given the information supplied by the premises, the conclusion does not tell something new.
These definitions are broad enough to fit both declaratives and imperatives.
The theory of imperatives that I will present gives a dynamic twist to the theory developed by Paul Portner (e.g. 2004), the basic idea being that an imperative sentence invites the addressee to update his or her plans for the future with the action described in the imperative.

Problem 2. Unmixable modalities

In all natural languages the possibilities to combine different modalities in one sentence are limited. It is easy to put a deontic modality in the scope of an epistemic modality, or an epistemic operator in the scope of an evidential expression, but it is impossible to do these things the other way around. Compare

- Maybe you should stop judging books by their cover.
- Clearly, he must be a spy.

with

- It ought to be case that he might be ill. (??)
- Maybe he is clearly a spy. (??)

[1] Actually, much of the work was done by my students Fabrice Nauze and Rosja Mastop.

How to explain this? Within the standard approach no explanation can be given. Actually, on the standard account (*locus classicus* is Kratzer, 1981), in which all modal expressions are treated as sentential operators that implicitly refer to some accessibility relation between possible worlds, it should be possible to mix all kinds of modalities in all possible ways. 'Maybe it is the case that it ought to be the case that ϕ' says that there is some epistemically accessible world w such that ϕ is true in all worlds that are ideal from the perspective of w. And 'It ought to be the case that maybe it is the case that ϕ' says that in all ideal worlds there is an epistemically accessible world in which ϕ is true. If there is nothing wrong with the first combination, what could be wrong with the second? The treatment of deontic modalities that I will propose makes clear why it is odd, if not impossible, to have an epistemic modality in the scope of a deontic modality. I will compare this explanation with the explanation given by Nuyts (2004), who looks at the problem from a cognitive-functional perspective.

REFERENCES

Paul Portner, 2004, 'The Semantics of Imperatives within a Theory of Clause Types', in: K. Watanabe and R. B. Young (eds.), *Proceedings of Semantics and Linguistic Theory 14*, Ithaca, NY: CLC Publications.

Angelika Kratzer, 1981, 'The notional category of modality', in: H-J. Eikmeyer and H. Rieser (eds.), *Words, Worlds, and Contexts – New Approaches in Word Semantics*, pp. 38-74, Berlin.

Jan Nuyts, 2004, *Over de (beperkte) combineerbaarheid van deontische, epistemische, en evidentiële uitdrukkingen in het Nederlands*. Antwerp papers in linguistics 108, 138 pp, Antwerp.

Saturday 19 November 2005, 9:00-13:00, fifth session:

2.5. Computation and applications: Knowledge dynamics in logic and Artificial Intelligence

Chair: Frank Veltman (University of Amsterdam)

A survey of dynamic epistemic logic

Wiebe van der Hoek

University of Liverpool, WiebevanderHoek@csc.liv.ac.uk

When giving an analysis of knowledge in multiagent systems, one needs a framework in which higher-order information and its dynamics can both be represented. Our work contributes to such a framework. It also fits in approaches that not only dynamize the epistemics, but also epistemize the dynamics: the actions that (groups of) agents perform are epistemic actions. Different agents may have different information about which action is taking place, including higher-order information. We demonstrate that such information changes require subtle descriptions. Our contribution is to provide a complete axiomatization for an action language, in which an action is interpreted as a relation between epistemic states (pointed models) and sets of epistemic states. The applicability of the framework is found in every context where multiagent strategic decision making is at stake, and already demonstrated in game-like scenarios such as Cluedo and card games.

A unified setting for inference and decision: An argumentation-based approach

Leila Amgoud

IRIT-CNRS Toulouse, amgoud@irit.fr

Decision making and inference have been studied for a long time separately, and have been considered as two distinct problems. The basic idea behind inference is to make ``safe" conclusions from a set of premises, whereas the decision making problem consists of selecting the ``best" decision among different alternatives on the basis of the available information (the beliefs about the environment, the goals, etc.).

In this paper we argue that inference is part of a decision process. The basic idea is to infer from all the available information, the formulae which are ``correctly" supported, then to order the different decisions only on the basis of these formulas. Thus, a decision problem can be seen as a two steps process: i) inferring from inconsistency then ii) ordering the alternatives using any criterion among those defined in classical decision theory.

We propose a general *argumentation framework* in which the two problems are analyzed and handled. Argumentation is a reasoning model which follows the following steps:

1. constructing arguments (*in favor* of /*against* a ``statement") from bases,
2. defining the strengths of those arguments,
3. determining the different conflicts between the arguments,
4. evaluating the acceptability of the different arguments, and
5. concluding.

Argumentation may also be considered as a different method for handling uncertainty. The basic idea behind argumentation is that it should be possible to say more about the certainty of a particular fact than the certainty quantified with a degree in [0, 1]. In particular, it should be possible to assess the reason why a fact holds, in the form of arguments, and combine these arguments to evaluate the certainty. Indeed, the process of combination may be viewed as a kind of reasoning about arguments themselves in order to determine the most acceptable of them.

Such an approach has indeed some obvious benefits, in particular, it is more acute with the way humans often deliberate and finally make a choice. Indeed, different arguments in favour of and against each decision are constructed, and pairs of decisions are compared on the basis of the quality of those arguments. Three kinds of arguments are distinguished: *epistemic* arguments that support beliefs, *recommending* arguments and *decision* arguments that support decisions. Different criteria criteria for evaluating the strength of arguments are given and criteria for comparing arguments are also discussed. Moreover, three categories of criteria for comparing decisions are proposed: i) *unipolar* criteria that take into account only one kind of arguments when comparing pairs of decision, i.e only arguments in favour of the two decisions, or only the arguments agaisnt the decisions are considered. ii) *bipolar* citeria where both arguments in favour of and against each decision are taken into account, iii) *nonpolar* criteria which consist of aggregating the different arguments of each decision into a single argument, and then to compare decisions on the baisis of the quality of their aggregated arguments.

Another feature of the proposed framework is that it extends classical work on decision theory in the sense that the hypothesis that the information about the environment is coherent is no longer required by this general framework. Moreover, the framework is general enough to capture different kinds of decision problems, namely, decision under uncertainty, multiple criteria decision and rule-based decision.

Possibilistic logic and knowledge dynamics

Henri Prade

IRIT-CNRS Toulouse, prade@irit.fr

The talk will provide an overview of the applications of possibilistic logic to various knowledge and preference dynamics problems. Possibilistic logic is an extension of classical logic, where logical formulas are associated with priority levels belonging to a linearly ordered scale. These priority levels can be thought as levels of entrenchment of the formulas. Formulas may represent pieces of knowledge entertained by agents, as well as goals pursued by agents. An important feature of possibilistic logic is its capability to handle inconsistency. Indeed a level of inconsistency is associated with any possibilistic logic base. It determines the part of the logic base, made of formulas with high priority levels, which is safe from any inconsistency. Extensions of possibilistic logic, having an argumentative flavor, can also take advantage of formulas having a level of priority below the inconsistency level. Besides, a bipolar representation framework enables us to distinguish between negative and positive information. The first type of information corresponds to standard logical information that restricts the set of possible interpretations (which is the exact complement of negative information stating what is impossible). Then the more information, the more restrictive the set of interpretations obtained as the intersection of the elementary restrictions associated with each granule of information. Obviously, this set becomes empty in case of inconsistency. Positive information refers to observed states of fact that can be accumulated in a disjunctive manner. Interestingly enough possibilistic logic bases, which induce a complete preorder on the set of interpretations at the semantic level, can be also represented under the form of a graphical Bayesian-like structure, or as a set of conditionals.

The presentation will first consider knowledge change problems referring to a static world, such as belief revision with certain or uncertain inputs, multiple source information merging, in a uni-polar, or a bipolar setting. In this latter case, consistency should not be only maintained for the restrictive part of the information, but also with respect to positive information (since what is feasible because observed should not be made impossible by the other part of the information). Besides, in case of information merging, the possibilistic framework enables us both i) to look for compromises in case of sets of conflicting goals of a group of agents, or ii) to find out what is plausibly true in the real world in case of inconsistent pieces of knowledge, using different merging operators. This applies both to classical and to possibilistic knowledge or preference bases. Belief and goal revision will be also considered in case of several interacting agents, such as in a negotiation process. Lastly, problems referring to a dynamical world will be discussed. This will include updating, prediction and post-diction, Kalman-like filtering problems. Kalman-like filtering corresponds to a type of updating involving a prediction step followed by a revision step. Updating operations based on imaging in the sense of Lewis appears as a particular case of Kalman-like filtering. All the considered operations on possibilistic logic bases can be performed at the syntactic level, in agreement with the possibilistic semantics.

Distributed vs. centralized belief revision

Aldo Franco Dragoni

Technical University of Marche, a.f.dragoni@univpm.it

Some time ago, an eminent biologist, asked whether there will ever be, on earth, organisms more complex than human brain, answered that such an astonishing super-brain already exists and it is the society of the minds interacting on the planet. Indeed, the claim that thinking should be regarded as a social phenomenon is a well known theory of cognitive psychology. The social view of cognition could help to understand how new scientific theories emerge and old philosophical currents disappear, but it could also explain simple collective phenomena as: how was it possible that almost an entire people believed that Iraq held weapons of mass destruction even if none of its members ever saw them (presumably)?

No doubt that most of our opinions are not elaborated from the others but, simply, adopted. So the main questions seems:

1 how do we choose (possibly unconsciously) the sources of information to believe in, among the many possible conflicting ones?

2 how do individuals criteria for evaluating the others reliability affect the global social cognitive behavior?

Question 1 is common to all of us, while the second is more academic and, probably, could interest few people but sociologists and computer scientists. In fact, both questions could have a descriptive answer (how people decide) and a normative answer (how people should decide in order to improve performances, satisfy requirements or reach goals). The latter kind of answer could interest software engineers building distributed problem-solving systems in which each node is affected by some degree of incompetence. So, we are focusing on a very specific problem of collective intelligence: the distributed elicitation of knowledge, i.e., how is it possible that from a variety of inferential schemas and judging capabilities, from different opinions and dogmas, from distinct perspectives and opposite point of views, after a continual interaction, a more uniform (if not unique) vision of the world emerges? Under which extraordinary circumstances the emergent representation of the world results a correct one?

Our hope is that of capturing some successful mechanisms in order to replicate them in a world of intelligent highly-engineered (programmed or trained) interacting cognitive agents.

As anticipated, a central question is: how should each agent ascribe a relative degree of reliability to any member of its group? Each agent should even be able to evaluate its own reliability. Even when agents are supposed not to lie, it is important to define methods for assessing each member's relative degree of reliability.

We limited our attention to groups in which this ascription is performed under liberal policies, i.e., each one is permitted to stand on its own opinions, evaluating him/herself and the others on the basis of the reciprocal experience and acquaintance.

From a global normative perspective we distinguish two desiderata:

1. convergence: are there local cognitive strategies which favor the convergence of the opinions (independently from their correctness w.r.t the real world)?
2. correctness: which of these local cognitive strategies favor also the correctness of the opinions?

Convergence is not trivial; think, for instance, a criterion such as: always believe the last information to arrive among conflicting information; it does not guarantee convergence under any policy of communication.

These criteria are almost conscious to humans. On the contrary, the way we form our opinions, from directly perceived material and from information received from the others, seems to be partially unconscious. Perhaps, one makes at least the following kinds of check after the incoming of a new information from an external source:

1. although I was not aware of it, is this new information a logical consequence of my beliefs about the world?
2. although I was not aware of it, is this information in accordance with my beliefs about the world?
3. although I was not aware of it, is this information in accordance with my direct experience of the world?
4. how many people believe it?
5. how much reliable are those people?
6. what is the source's goal when saying that to me?
7. how much relevant is that to me?

1 represents a mere confirmation, while 2 and 3 may yield a consistent expansion of the agent's knowledge. People who stress the importance of 3 are confident in themselves. Those who prefer 4 are rather conformist, while check 5 is preferred by suspicious people; noticeable, 5 links the

problem of establishing the credibility of the various pieces of information to the problem of evaluating the reliability of their respective sources.

Perhaps 6 should be the most important to humans; we are continually addressed by commercial advertisments whose obvious goals should be taken into account when evaluating their truthfulness. However, we avoids the problem of goals recognition and goals treatment in order to evaluate the credibility of an information because the artificial agents we have in mind have no hidden agendas and most of them only have inherent aims, furthermore we experienced the need to simplify the scenario to reach some conclusions, even if they will be partial. We also completely escaped from dealing with relevance (7); we simply assumed that all the pieces of information running through the network were relevant for each one of its nodes.

Question 2 before could be rephrased as follows: if all the individuals adopt the same local criteria to evaluate an incoming information, how is the global process of knowledge elicitation affected?

For instance, if all the individuals only adopt check 2, then we would expect little gains from the interaction, each staying in his native degree of correctness. On the other hand, if all the individuals exclusively adopt the check 4, with no regard toward his own and the others' competence, then we would expect a global flattening to the medium degree of correctness of the agency. Perhaps we should adopt a reasonable mix of these criteria. Unfortunately, these criteria are too vague to be studied on a statistical simulation basis. We need more precise rules. Specifically, we want to understand what happens to the emergent group's opinions when its members cognitively act in certain manner and all obey a common communication policy.

The group's cognitive performance could be evaluated under different perspectives:

1. local perspective: by measuring each individual s derived benefit from having been part of the group
2. global perspective: by comparing the group's global opinions under different strategies of belief revision and different policies of communication

We tried to make these evaluations by means of simulation. Results will be reported in the talk.

Agents changing their minds about the others’ minds: Belief revision in multi-agent systems

Rineke Verbrugge

University of Groningen, rineke@ai.rug.nl

Introduction

I will report on research on group attitudes and communication in multi-agent systems, done in collaboration with Barbara Dunin-Keplicz, as well as research on cognitive limitations on reasoning about other agents with Petra Hendriks, Irene Kramer, and Lisette Mol.

Changing beliefs by reasoning about others

In everyday situations it is extremely important to reason about other people's minds: you need to reason about their knowledge in order to interpret what they say and to construct your own utterances so as to be understood; you need to reason about their beliefs and intentions in order to negotiate with them; for many situations you even need to reason about their beliefs about your beliefs about them, and about even higher-order beliefs.

Formal models of human reasoning, such as those in epistemic logic and game theory, assume that humans can faultlessly reason about other people's knowledge and about common knowledge, for example in card games such as happy families (Hoek&Verbrugge 2002). However, recent research in cognitive psychology reveals that adults do not always correctly use their theory of what others know in concrete situations (Keysar 2003, Hedden & Zhang 2002).

In Keysar's experiments, some adult subjects could not correctly reason in a practical situation about another person's lack of knowledge (first-order theory of mind reasoning) (Keysar 2003).

Hedden and Zhang, when describing their experiments involving a sequence of dyadic games, suggested that players generally began with first-order reasoning. When playing against first-order co-players, some began to use second-order reasoning, but most of them remained on the first level (Hedden & Zhang 2002).
In recent experiments by AI Master's student Lisette Mol (Mol 2005-1, Mol 2005-2), it turns out that humans *can* learn to play a version of symmetric Mastermind involving natural language utterances such as "some colors are right". After mastering the first task, namely to play the game according to its rules, many of them learn to perform a second task, namely to develop a winning strategy for the game by using a higher-order theory of mind: "Which sentences reveal the least information while still being true?" "What does the opponent think I am trying to make him think?"
In the talk, we will discuss the results of these experiments, that comply with other researchers' findings that theory of mind of higher than second order is seldom used in practical situations.

Cognitive limitations and common knowledge
If even limited orders of theory of mind present such difficulties for humans, it seems that reasoning about common knowledge, which apparently involves an infinitude of levels, is impossible. From the time when the notion of common knowledge was first studied, there has been a puzzle about their establishment and assessment, the so-called *Mutual Knowledge Paradox*, most poignantly described in (Clark & Marshall, 1981). How can it be that to check whether one makes a felicitous reference when saying "Have you seen the movie showing at the Roxy tonight", one has to check an infinitude of facts about reciprocal knowledge, but people seem to do this in a finite, indeed short, time?

Changing group beliefs by communication in multi-agent systems
Notions of knowledge about others and group knowledge also play an important role in *multi-agent systems*, where a number of computational agents work together in order to solve a problem that they cannot solve on their own. Indeed common knowledge is seen as the basis of coordination among agents. Halpern and Moses proved that common knowledge of certain facts is on the one hand necessary for coordination in well-known standard examples, while on the other hand, common knowledge cannot be established by communication if there is any uncertainty about the communication channel (Fagin, 1995).

Belief and common belief
In practice in multi-agent systems, agents often make do with belief instead of knowledge for the following reasons. First, in multi-agent systems, perception provides the main background for beliefs. In a dynamic, unpredictable environment, the natural limits of perception may give rise to false beliefs or to beliefs that, while true, still cannot be fully justified by the agent. Second, communication channels may be of uncertain quality, so that even if a trustworthy sender knows a certain fact, the receiver may only believe it.
Common belief of p is the notion of group belief which is constructed in a similar way as common knowledge: everyone believes p, everyone believes that everyone believes p, and so on, ad infinitum. Note that, in contrast to common knowledge, which is always sure, common belief need not be truthful, thus in some situations it may be a common illusion. (See (Fagin, 1995; Hoek & Verbrugge, 2002) for more about logics for common knowledge and belief).

Problems in creating common knowledge
Halpern and Moses (Halpern & Moses 1984) proved a surprising result in the eighties: under some very natural assumptions, namely that processors do not change their local states simultaneously, common knowledge does not increase over a run (sequence of time steps) in a distributed system. The well-known example of the two generals who do not manage to reach common knowledge about the time of attack, even if a messenger brings any number of acknowledgements back and

forth, is an example of this result. If there is any uncertainty about the messenger making it to the other general, even about whether he may be delayed, common knowledge cannot be reached. In multi-agent systems, there is almost always uncertainty about messages reaching the other party. In the talk, we will discuss problems and solutions related to changing a group's mental state in a multi-agent system.

Saturday 19 November 2005, 15:00-19:30, sixth session:

2.6. Changes in view: Future developments and priorities in the study of knowledge dynamics

Chairs:
Johan van Benthem (University of Amsterdam / Stanford University)
Cristiano Castelfranchi (ISTC-CNR Roma / University of Siena)
Andreas Herzig (IRIT-CNRS, Toulouse))
Boicho Kokinov (New Bulgarian University)
Elizabeth Robinson (University of Warwick)
Hans Rott (University of Regensburg)

The concluding session of the workshop begun with an open debate among all the participants, supervised by the organizing committee, i.e. Cristiano Castelfranchi, Boicho Kokinov and Fabio Paglieri. The aim was to envision future collaborative initiatives by first focusing on unresolved issues, top priorities, and critical challenges in the field of knowledge dynamics. After approximately 60 minutes of debate, the following two needs emerged as the most urgent and all-important:

- concerning scientific focus, the topic of *trust and its dynamics* was mentioned several times as a crucial issue, one on which several different disciplines could and should strive to provide a more comprehensive and integrated account;
- as for interdisciplinary background, the necessity of improving *cross-fertilization and integration* of different approaches to knowledge dynamics was stressed unanimously. All participants agreed that this workshop should serve as first impulse to realize more ambitious networking activities among the researchers and the institutions committed to the study of knowledge dynamics, both in Europe and in other countries.

Parallel round-tables on research priorities and scientific challenges

To better capitalize the insights emerged during the initial debate, the workshop participants were subsequently split in three different round-tables, each of them focused on one of the following topics:

- *The psychology of knowledge dynamics* (chairs: B. Kokinov, E. Robinson)
- *The technologies of knowledge dynamics* (chairs: A. Herzig, H. Rott)
- *Knowledge dynamics in social interaction* (chairs: J. van Benthem, C. Castelfranchi)

These round-tables run in parallel for approximately 60 minutes. Each group was coordinated by two of the session chairs, who supervised the debate and make sure that the intended objectives were properly addressed. Each round-table discussed, with reference to its specific topic, the following features:

1. Most significant and reliable results already achieved and available in that domain

2. Open problems, unresolved issues, and most urgent scientific challenges
3. Pros and cons of available methodologies in dealing with such problems (including perspective for integration of different methodologies and development of novel approaches)
4. Opportunities for interdisciplinary research (including an analysis of possible risks and pitfalls)
5. Expected societal impact of these lines of research (both short-term and in the long run)

At the end of these consultations, each work-group appointed a spoke-person among its members, who shortly reported on the round-table results at the beginning of the next session.

Plenary discussion of possible follow-ups and future initiatives

The final session of the workshop was aimed to pool together the main themes surfaced in the previous days, to start finalizing scientific follow-ups and future research initiatives among the participants. To this purpose, the works were organized as follows:

- 17:30-18:00 The spoke-persons from each round-table made short reports on the results emerged from each work-group, with special emphasis on the most challenging research issues, and the most indicated methodologies to deal with them
- 18:00-18:30 A short survey of the available means to foster future collaboration among workshop participants was presented, including: relevant calls for projects, research networks, mobility programmes, organization of future events, bilateral agreements. First a brief survey of ESF activities was presented by Prof. Bengt Hansson, Lund University; then a complementary survey was outlined by Dr. Fabio Paglieri, covering other EU-financed initiatives.
- 18:30-19:30 Finally, the results of the round-tables and the potentialities for follows-ups were assessed in an open debate among workshop participants, aimed to finalize some concrete proposals for future cooperation, and to set a first tentative schedule to realize such cooperation in practice. This concluding discussion resulted in the initiatives listed in section 3 of this report.

3. Assessment of the results, contribution to the future direction of the field

The scientific results of the workshop were fully satisfactory, both for the organizers and for the participants, who unanimously declared that the event had been extremely successful, insightful, and fruitful for their own lines of research. Within the broader field of knowledge dynamics, one specific topic emerged as especially urgent and promising, at the convergence of several different disciplinary approaches: the nature of the concept of *trust and its dynamics*, as the most fundamental building block for generating reliable knowledge from raw information. This theme proved to be crucial for several different approaches to knowledge dynamics, both in the humanities (e.g. philosophy, psychology, anthropology), in the social sciences (e.g. economics, sociology), and in applied disciplines (e.g. computer science). All the scholars involved in the workshop, regardless of their own disciplinary background, agreed that this topic will hence require *major interdisciplinary efforts* to be adequately understood and modelled, and to develop effective applications of trust technologies and practices. In fact, as described below, future research initiatives, jointly planned by the workshop participants and their institutions, aim to tackle this issue via well-focused multi-disciplinary networks and projects, that are expected to provide continuity and further momentum to the scientific achievement of this exploratory event.

Future joint research initiatives

(1) Joint proposal for a Marie Curie Research Training Network on the topic "Knowledge Dynamics", as soon as the next call is out (possibly next summer, with deadline in autumn). This will serve to finance mobility of researchers (especially young ones, but not only) among the participants' institutions, initiating and developing more in-depth collaborations.

(2) Joint proposal of the topic "Nature and Dynamics of Trust" as the theme for one of the future EUROCORES Programmes financed by the ESF within the Humanities area. We expect the next call to be out next Spring - therefore, if our proposal will be successful, we may expect to launch a call for Collaborative Research Projects on this topic during the year 2007 (for more information on this scheme, consult http://www.esf.org).

(3) In addition, we look forward to and intend to promote more narrowly focused initiatives resulting from this workshop, involving 2 or more researchers and their respective research teams (research projects, bilateral agreements, etc.). In fact, the general discussion on the last day showed that both top-down and bottom-up processes are to be encouraged, in order to improve and broaden our mutual cooperation in the future on these exciting topics.

Future events and meetings

Although the participants agreed that in the immediate future it may be more profitable to focus on research initiatives, they also consider valuable to ensure resources for future meetings, both larger and smaller than this Exploratory Workshop. To this purpose:

- for smaller meetings, aimed to focus on more specific issues and to produce in-depth technical discussion, the impulse will come from single institutions, that, by using either their own funds or national financial support, could host small-sized events involving only a fraction of our larger network (a good example of such gathering was realized at the ILLC-UvA in Amsterdam last year, see http://www.unisi.it/ricerca/dip/fil_sc_soc/dot-sc/belrev.html);
- for larger meetings, involving much more people and tailored to foster interdisciplinary debate on several topics, we are considering to apply for EU funds, e.g. submitting a proposal for the next call in the Marie Curie Conferences and Training Courses (expected publication: 18 January 2006 // expected deadline: 17 May 2006). Basically, this scheme, if successful, would ensure us funds for at least 4 events spanned over 4 years, possibly including both large conferences and intensive training courses for young researchers (e.g. summer schools). More information can be found at http://europa.eu.int/comm/research/fp6/mariecurie-actions/action/courses_en.html

Publications

We plan to realize an edited volume on "The dynamics of knowledge", composed by the revised and expanded versions of most of the workshop contributions, that should appear in late 2006 / early 2007. As for the publisher, we are considering several options, but at this time the more likely candidate appears to be Cambridge Scholars Press, UK.

In the meantime, more narrowly focused editorial initiatives are well under way, involving several of the workshop participants: among other projects, Fabio Paglieri is editing a special issue of the international journal *Synthese: Knowledge, Rationality and Action* on "Changing minds: Cognitive, computational, and logical approaches to belief dynamics", to appear in late 2006, while Johan van Benthem (in collaboration with Helen and Wilfrid Hodges) is editing a special issue of *Topoi: An international review of philosophy* on "Logic and Cognition", to appear in early 2007. Most of the workshop participants, together with other top scholars in the field, are contributing to either one of the two special issues. Other similar projects are being pursued by other scholars involved in this workshop, and they will help to further strengthen the links of mutual cooperation and scientific exchange built on this occasion.

4. Final programme

THURSDAY 17 NOVEMBER 2005

8.30 Opening remarks

9.00-13.00
Short-term dynamics of knowledge: Cognitive and computational models of belief change
Chair: Elizabeth Robinson (University of Warwick)

Boicho Kokinov (New Bulgarian University)
Context-sensitivity of human cognition: Fast short-term restructuring and adaptation of the cognitive system based on what is anticipated to be relevant

Robert French (University of Burgundy)
Fluidly represent the world: Way, way harder than you think

Cristiano Castelfranchi (ISTC-CNR, Roma)
From knowledge to action: Reason-based belief dynamics and belief-based goal dynamics

Coffee break

Natalie Gold (Duke University)
Belief dynamics, framing effects and decision-making

Fabio Paglieri (University of Siena)
Data, beliefs, and acceptances: Ontology and dynamics of doxastic states

13.00 Lunch

15.00-19.15
Dynamics of knowledge in childhood: Cognitive and developmental perspectives
Chair: Boicho Kokinov (New Bulgarian University)

Elizabeth Robinson (University of Warwick)
Belief formation and change in human development

Stella Vosniadou (University of Athens)
Knowledge acquisition and conceptual change in childhood

Han van der Maas (University of Amsterdam)
Cognitive development: The balance scale task

Coffee break

Vittorio Girotto (IUAV, University of Venezia), Michel Gonzalez (CNRS / University of Provence)
Young children's intuitions about posterior probability

Kristien Dieussaert, Deborah Vansteenwegen, An Van Assche (University of Leuven)
On the stability of instruction and experience based beliefs

20.00 Dinner

FRIDAY 18 NOVEMBER 2005

9.00-13.00
Evolutionary and socio-cognitive dynamics of knowledge
Chair: Cristiano Castelfranchi (ISTC-CNR, Roma)

Peter Gärdenfors (Lund University)
The evolution of a theory of mind, common knowledge and cooperation

Maurice Grinberg (New Bulgarian University)
Evolution of conceptual maps as a result of learning

Antoine Billot (PSE-ENPC, CORE), Jean-Christophe Vergnaud (CNRS-Eureqa), Bernard Walliser (PSE-ENPC, EHESS)
Multi-player belief revision and information value in games

Coffee break

Brian Loasby (University of Stirling)
Knowledge in economics

Paolo Turrini, Mario Paolucci, Rosaria Conte (ISTC-CNR, Roma)
Word of mouth: The added value of beliefs' dynamics

13.00 Lunch

15.00-19.15
Formal models of knowledge dynamics: Comparison with social and cognitive theories
Chair: Peter Gärdenfors (Lund University)

Johan van Benthem (University of Amsterdam / Stanford University)
Cognition as interaction

Andreas Herzig, Benoit Gaudou, Dominique Longin (IRIT-CNRS, Toulouse)
Analyzing communication in a logic of grounding, belief and intention

Hans Rott (University of Regenburg)
Shifting priorities: Simple representations for twenty-four iterated theory change operators

Coffee break

Krister Segerberg (Uppsala University), Hannes Leitgeb (University of Bristol)
Dynamic doxastic logic: Why, whether, how

Frank Veltman (University of Amsterdam)
Mixed moods and unmixable modalities. Modelling beliefs and intentions

19.30 Dinner in Siena

SATURDAY 19 NOVEMBER 2005

9.00-13.00
Computation and applications: Knowledge dynamics in logic and Artificial Intelligence
Chair: Frank Veltman (University of Amsterdam)

Wiebe van der Hoek (University of Liverpool)
A survey of dynamic epistemic logic

Leila Amgoud (IRIT-CNRS, Toulouse)
A unified setting for inference and decision: An argumentation-based approach

Henri Prade (IRIT-CNRS, Toulouse)
Possibilistic logic and knowledge dynamics

Coffee break

Aldo Franco Dragoni (Technical University of Marche)
Distributed vs. centralized belief revision

Rineke Verbrugge (University of Groningen)
Agents changing their minds about the others' minds: Belief revision in multi-agent systems

13.00 Lunch

15.00-19.30
Changes in view: future developments and priorities in the study of knowledge dynamics
Chairs: J. van Benthem, C. Castelfranchi, A. Herzig, B. Kokinov, L. Robinson, H. Rott

Open debate among workshop participants, oriented towards the definition of future scenarios and innovative approaches to knowledge dynamics, in order to envision the research agenda in this field for the next decade and to foster further cooperation among the participants (joint projects, exchange programs, excellence networks, etc.)

19.30 End of works

20.00 Farewell dinner

5. Final list of participants with contact details

The total number of participants to the workshop was 34. Their names, institutional affiliations and contact details are listed below in two separate lists, the first for the invited speakers, the second for all the other participants.

Invited speakers (in alphabetical order)			
Amgoud, Leila	amgoud@irit.fr	IRIT-CNRS 118, route de Narbonne 31062 Toulouse Cedex, France	Tel: (00 33) 5 61 55 64 19 Fax: (00 33) 5 61 55 62 39
Castelfranchi, Cristiano	cristiano.castelfranchi@istc.cnr.it	ISTC-CNR Via S. Martino della Battaglia 44 00185 Roma, Italy	Tel: 0039 06 44595283 Fax: 0039 06 44595243
Conte, Rosaria	rosaria.conte@istc.cnr.it	ISTC-CNR Via S. Martino della Battaglia 44 00185 Roma, Italy	Tel: 0039 06 44595290 Fax: 0039 06 44595243
Dieussaert, Kristien	kristien.dieussaert@psy.kuleuven.ac.be	KULeuven Laboratory of Experimental Psychology Tiensestraat 102 B- 3000 Leuven, Belgium	
Dragoni, Aldo Franco	a.f.dragoni@univpm.it	Dip. di Elettronica, Intelligenza Artificiale e Telecomunicazioni Università Politecnica delle Marche Via Brecce Bianche 60131, Ancona, Italy	Tel: +39 071 2204390 Fax: +39 071 2204474
French, Robert	rfrench@ulg.ac.be, robert.french@u-bourgogne.fr	LEAD-CNRS UMR 5022 Pôle AAFE - Esplanade Erasme University of Burgundy 21065 Dijon, France	Tel: +33 [0]3.80.39.90.65 Fax: +33 [0]3.80.39.57.67
Gärdenfors, Peter	Peter.gardenfors@lucs.lu.se	Lund University Cognitive Science Kungshuset, Lundagård SE-222 22 LUND, Sweden	Phone: +46 46 222 48 17 Fax: +46 46 222 44 24
Girotto, Vittorio	girotto@mail.univ.trieste.it	Università IUAV di Venezia Dip. delle Arti e del Disegno Industriale Dorsoduro 2206 - ex Convento delle Terese 30123 Venezia, Italy	Fax 0039 041 257 1392
Gold, Natalie	goldnk@duke.edu	Dept. Of Philosophy Duke University 201 West Duke Building Durham, North Carolina 27708-0743, USA	Phone: 919-660-3065
Grinberg, Maurice	mgrinberg@nbu.bg	Central and East European Center for Cognitive Science New Bulgarian University 21 Montevideo Blvd. Sofia, 1618, Bulgaria	Phone: + 359 2 811 0401 Fax: + 359 2 811 0421
Herzig, Andreas	Andreas.Herzig@irit.fr	IRIT-CNRS 118, route de Narbonne 31062 Toulouse Cedex, France	Tel: +33 56155-8123 Fax: +33 56155-8898
Kokinov, Boicho	bkokinov@nbu.bg	Central and East European Center for Cognitive Science New Bulgarian University 21 Montevideo Blvd. Sofia, 1618, Bulgaria	Tel.: (+359) 2-9571876 Fax: (+359) 2-558262 or 565037
Loasby, Brian	b.j.loasby@stir.ac.uk	3B79 Cottrell Bldg. Dept. of Economics	Phone: +44-(0)1786-46-7489

		University of Stirling Stirling FK9 4LA, UK	Fax: +44-(0)1786-46-7469
Paglieri, Fabio	paglieri@media.unisi.it	ISTC-CNR Via S. Martino della Battaglia 44 00185 Roma, Italy	Mobile: 0039 348 3542376 Tel: 0039 06 44595310 Fax: 0039 06 44595243
Prade, Henri	prade@irit.fr	IRIT-CNRS 118, route de Narbonne 31062 Toulouse Cedex, France	Tel : 05 61 55 65 79
Robinson, Elizabeth	E.J.Robinson@Bham.ac.uk	Department of Psychology University of Warwick Coventry CV4 7AL, UK	Tel: (024) 761 50039 Fax: (024) 765 24225
Rott, Hans	hans.rott@psk.uni-regensburg.de	Institut für Philosophie Universität Regensburg 93040 Regensburg, Germany	Tel: +49 [0]941 943 3660 Fax: +49 [0]941 943 1985
Segerberg, Krister	krister.segerberg@filosofi.uu.se	Filosofiska institutionen Uppsala Universitet Box 627 751 26 Uppsala, Sweden	Tel: 018-471 73 52 Fax: 018-471 73 70
Van Benthem, Johan	johan@science.uva.nl	ILLC, University of Amsterdam Plantage Muidergracht 24 1018 TV Amsterdam, The Netherlands	Phone: +31 20 525-6051 Fax: +31 20 525-5206
Van der Hoek, Wiebe	wiebe@csc.liv.ac.uk	Dept. of Computer Science University of Liverpool Liverpool L69 7ZF, UK	Tel (+44 151) 794 3672/7480 Fax (+44 151) 794 3715
Van der Maas, Han	h.l.j.vandermaas@uva.nl	Dept. of Psychology University of Amsterdam Roetersstraat 15 1018 WB Amsterdam, The Netherlands	Phone +31-205256678 Fax +31-206390279
Veltman, Frank	veltman@illc.uva.nl	ILLC University of Amsterdam Nieuwe Doelenstraat 15 1012 CP Amsterdam, The Netherlands	Phone: +31 20 5254564
Verbrugge, Rineke	l.c.verbrugge@ai.rug.nl	University of Groningen Artificial Intelligence Grote Kruisstraat 2/1 9712 TS Groningen, The Netherlands	Phone: +31 50 363 6334
Vosniadou, Stella	svosniad@phs.uoa.gr	National and kapodistrian University of Athens, Dept. of Philosophy and History of Science, Panepistimioupolis, 157 71, Athens, Greece	Tel: +30-210-7275507 Fax: +30-210-7275504
Walliser, Bernard	walliser@enpc.fr	CERAS Ecole Nationale des Ponts et Chaussées 28 rue des Saints Pères 75007 Paris, France	Tel: 01 44 58 28 72 Fax : 01 44 58 28 80

Registered attendees (in alphabetical order)

DIETRICH ALBERT
DEPARTMENT OF PSYCHOLOGY, UNIVERSITY OF GRAZ
UNIVERSITAETSPLATZ 2, A 8010 GRAZ, AUSTRIA
Phone +43 316 380 5118
Fax +43 316 380 9806
E-mail DIETRICH.ALBERT@UNI-GRAZ.AT

LUIGI BATTEZZATI
POLITECNICO DI MILANO, DIPARTIMENTO INGEGNERIA GESTIONALE
VIA GIUSEPPE COLOMBO 40, MILANO, ITALY
E-mail LUIGI.BATTEZZATI@POLIMI.IT

CLAUDIA CASADIO
UNIVERSITÀ DEGLI STUDI DI CHIETI-PESCARA
VIA DEI VESTINI, CAMPUS UNIVERSITARIO, 66100 CHIETI
Phone 0871 3556583
Fax 0871 552452
E-mail CASADIO@UNICH.IT

PAUL GOCHET
DEPT OF PHILOSOPHY, UNIVERSITY OF LIÈGE
32 PLACE DU XX AOÛT, 4000 LIÈGE, BELGIUM
Phone 32 2 7330404
E-mail PGOCHET@ULG.AC.BE

JAN HEYLEN
CATHOLIC UNIVERSITY OF LEUVEN
NAAMSESTRAAT 22, 3000 LEUVEN, BELGIUM
Phone 32496567493
E-mail JAN.HEYLEN@HIW.KULEUVEN.BE

HANNES LEITGEB
DEPT. OF PHILOSOPHY, UNIVERSITY OF SALZBURG
FRANZISKANERGASSE 1, A-5020 SALZBURG, AUSTRIA
Phone 0043 662 8044 4084
E-mail HANNES.LEITGEB@SBG.AC.AT

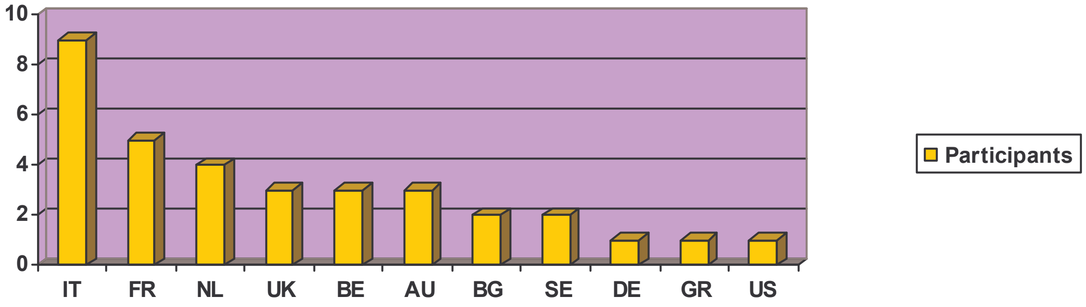
PAOLO TURRINI
UNIVERSITY OF SIENA
BANCHI DI SOTTO 55, 53100 SIENA, ITALY
Phone +39 3472814666
E-mail TURRINIPA@TISCALI.IT

MARKUS VINCZE
TECHNICAL UNIVERSITY OF VIENNA
GUSSHAUSSTR. 27/376, 1040 VIENNA, AUSTRIA
Phone 1,94626E-07
Fax 1,9444E-07
E-mail VINCZE@ACIN.TUWIEN.AC.AT

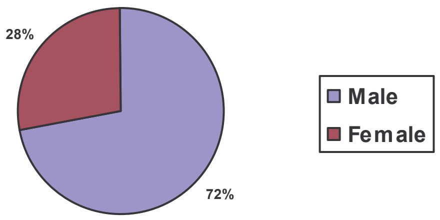
CHIARA ZINI
UNIVERSITY OF TRENTO
VIA MATTEO DAL BEN 5/B, 38068 TRENTO, ITALY
Phone +39 328 4214184
Fax +39 0464/483554
E-mail ZINI@FORM.UNITN.IT

6. Statistical information on participants

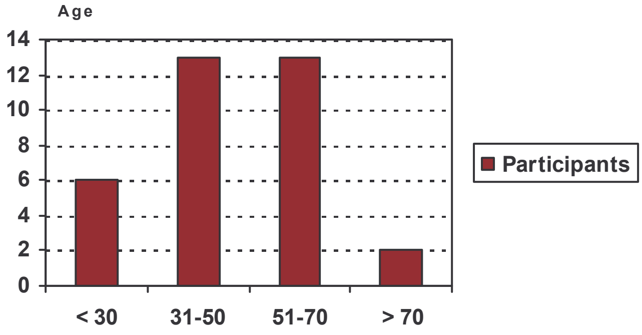
Eleven different countries were represented among our 34 participants, as summarized below: Italy (9), France (5), the Netherlands (4), United Kingdom (3), Belgium (3), Austria (3), Bulgaria (2), Sweden (2), Deutschland (1), Greece (1), United States (1).



As for gender distribution, 25 of the workshop participants were male, 9 were female (see below).



Age distribution was rather widespread, as detailed below: as for the active involvement of early stage researchers in the workshop, it should be notice that 17.6% of participants were below the age of 30, while 55.9% were below the age of 50.


Age
< 30
> 70
Participants